The Journal of Parapsychology

A SCIENTIFIC QUARTERLY DEALING WITH EXTRA-SENSORY PERCEPTION AND RELATED TOPICS

Volume 6 DECEMBER, 1942 Number 4

Contents

PAGE

Editorial: Progress of Parapsychology as a University Study . 237

An Experiment in Precognition Using Dice 243
J. L. WOODRUFF and J. B. RHINE

Some of the Convergences Which Initiated Parapsychology . 263
GERALD HEARD

The Role Mathematics Has Played in ESP Research . . . 268
LT. J. A. GREENWOOD, U. S. N. R.

A Single Subject in a Variety of ESP Test Conditions . . . 284
WILLIAM RUSSELL, U. S. Army, and J. B. RHINE

Glossary 312

DUKE UNIVERSITY PRESS
DURHAM, N. C.

The Journal *of* Parapsychology

Editors

J. B. RHINE C. E. STUART J. G. PRATT

J. A. GREENWOOD
Statistical Editor

DOROTHY H. POPE
Managing Editor

This journal is published on the fifteenth day of March, June, September, and December by Duke University Press, Durham, North Carolina.

Contributions submitted for publication and all editorial communications should be addressed to the Managing Editor, Dorothy H. Pope, College Station, Durham, North Carolina. Correspondence with the editors is advised before submitting articles other than reports of experimentation. Since it does not forward manuscripts by registered mail, Duke University Press cannot guarantee that they will not be lost in transit, and contributors are urged to keep copies of their papers. All contributions should be typewritten, double spaced. References should be given in the form adopted by THE JOURNAL.

Reprints may be ordered when the proof is returned. THE JOURNAL will bear one half the cost of reprints up to two hundred copies. Correspondence concerning subscriptions, change of address, back numbers, and other business communications should be addressed to the Parapsychology Laboratory, College Station, Durham, N. C.

The subscription price is $4.00 a year; single current numbers $1.00. The rate for back volumes is $5.00; for single numbers $1.25. Missing numbers will be supplied free when lost in the mails if written notice is given within one month of the date of issue. All remittances should be made payable to the Parapsychology Laboratory.

Entered as second class matter at the post office at Durham, N. C.

DUKE UNIVERSITY PRESS

The Journal *of*
Parapsychology

Volume 6 DECEMBER, 1942 Number 4

EDITORIAL:

PROGRESS OF PARAPSYCHOLOGY AS A UNIVERSITY STUDY

In our September editorial, we pointed to the successful emergence of hypnotism from its pioneer classification of sixty years ago under the heading of parapsychology (psychical research) and emphasized particularly the contrast between its acceptance in the scientific circles of university life and the slow recognition achieved by its long-associated partner, extra-sensory perception. We realize now, in retrospect, that this leaves a poor and incomplete impression of the real progress of the ESP research toward "full academic standing." We will, therefore, now attempt to supplement the previous account, considering some of the evidences of increased recognition, and omitting, for the moment, any unfavorable comparisons. (If we appear to equate the term "parapsychology" with the "ESP research" in this discussion, it is only because for the present, as well as for some years in the past, the latter has pretty completely monopolized the scene in parapsychology.)

In order to estimate progress of the kind we have in mind, it is necessary to find, distributed over the period to be surveyed, some criteria on which basic judgments can be made. Fortunately, we have such a basis in the addresses of three eminent psychologists, each of whom, during the last half-century, has made an appeal in behalf of parapsychology. These three spokesmen have been strategically situated for the use we intend to make of their remarks. For one thing, in point of time, they represent three fairly distinct "generations" in

the history of psychology. All three of them, too, have taken an active interest in the parapsychological field, although they are not classifiable as parapsychologists. Finally, two have served, and one is serving now, as President of the Society for Psychical Research. One was an American; one, a British-American; and the other is a Briton. All three are recognized as men of great breadth in their psychological interests and contributions. They are William James, William McDougall, and Robert Thouless.

From the writings of these men on the subject of parapsychology we select and quote from three essays or addresses dated 1896, 1922, and 1942 respectively. It will be evident, we think, that from the very estimates of the field which these three appeals clearly reflect, each for its own epoch, we can obtain three points of judgment from which a curve of progress may readily be drawn.

William James, in his essay, "What Psychical Research Has Accomplished," published in 1896, directed his remarks to the general public—of course, the more intellectual portion of the general public. His appeal was definitely one for the maintenance of the Society for Psychical Research and the furtherance of its efforts; only incidentally did he turn his attention to the university world, as when he said, "The young anthropologists and psychologists who will soon have full occupancy of the stage will feel how great a scientific scandal it has been to leave a great mass of human experience to take its chances between vague tradition and credulity, on the one hand, and dogmatic denial at long range on the other, with no body of persons extant who are willing and competent to study the matter with both patience and rigor." But all the more need, then, for the work of the Society! And today all who are familiar with the accomplishments of that Society will, we believe, agree that James did not exaggerate the importance of its mission.

But although James was himself a Harvard professor, there is nothing in his essay to suggest that he had any ground for hope, at that time, of establishing this branch of inquiry within the university world. And although he was a psychologist, neither his remarks nor his review of progress in the research reflect any interest on the part of any other members of that profession.

Looking forward to the year 1922, however, and to the address

by William McDougall, delivered at Clark University, entitled "Psychical Research as a University Study," we find that 26 years have greatly modified the status of that branch of inquiry; the mere fact that McDougall was speaking on a university lecture series devoted to the subject of psychical research is itself of some significance. It is still more meaningful, too, that at that time, or shortly thereafter, he was sponsoring an investigation of telepathy at Harvard University to which he had recently come. But the very pitch of expectation in his appeal reflects a confidence in the subject and in its future which could not have been held by any psychologist in the day of William James, for McDougall boldly asked that psychical research be admitted by universities to their program of research and study.

Speaking of telepathy, he remarked, "What more suitable task for a research department of a university can be conceived than the task of investigating such a problem?" And of the field in general he said, "Here the mind long disciplined in other branches of science may find the supreme test of its powers and its training, tests of impartial observation, of relevant selection, of sagacious induction and deduction, of resolute discounting of emotional bias and personal influence. Here, better than in any other field, it may learn to recognize its own limitations, limitations of knowledge, of power, of principle; and to recognize also the limitations of science and philosophy themselves, their inadequacy to give final answers to problems which mankind has long answered with ready-made formulae, handed down from the dim dawn of human reflection, and before which men now halt with burning desire for certainty or unsatisfied longing for more light."

When in 1933 the address was republished, Professor McDougall added in a postscript, "I am glad to be able to report that Duke University, which I have had the honor of serving since 1927, has followed the lead given by Clark University and has gone farther in the same direction. Psychical research has been not only prosecuted in the laboratory with very interesting results, but also furthered by a course of lectures. . . ."

But while McDougall asked for a great deal more than James, he still did not call for the general acceptance of any of the findings

of the field. He stated with great reserve some of his own personal convictions of the genuineness of certain occurrences of parapsychical character, but seemed to feel a general shakiness of the evidential case for most of the hypothetical phenomena.

Third in the series is the presidential address of Robert Thouless of Cambridge University which was delivered before the Society for Psychical Research on the first of May of this year. Thouless spoke as an experimental psychologist, and not only that, but as one who has experimented in problems of extra-sensory perception. This fact itself is significant, for it must be remembered that Thouless is not a parapsychologist. The point is that the problems of ESP and related hypotheses have penetrated the laboratories of experimental psychology today. Professor Thouless states, ". . . already there are probably more independent workers on the psi phenomenon (his term for ESP) than on any other psychological topic, and it is an encouraging fact that even odd and unexpected results are being confirmed by independent workers. The check of independent confirmation is working to a very considerable extent."

But the best criterion in the address of Professor Thouless of the state to which parapsychology has progressed since 1922 lies in the position he takes with regard to the establishment of extra-sensory perception. He calls clearly for an acceptance of this phenomenon: "By different methods, a number of workers have obtained under stringent experimental conditions positive results which cannot reasonably be attributed to chance or to experimental error. The work of obtaining these results has been laborious, and great credit is due to those who have undertaken it. A mere repetition of that work now would be a great waste of time. Let us get on to other problems, to be solved by other methods."

And he added as the final appeal of his address, "Finally I should like to suggest that the general result of our survey is that the present position of experimental research in these topics is a hopeful one. Existing researches have not merely proved the existence of the psi phenomena; they have also found out some odd and unexplained things about them that are a challenge to further research. Let us now give up the task of trying to prove again to the satisfaction of the sceptical that the psi effect really exists, and try instead to devote

ourselves to the task of finding out all we can about it. With fuller knowledge of its nature, the difficulties of believing in its existence may appear less formidable than they do now."

This is a far-advanced position from those represented in the earlier expressions quoted. It tells us, as we interpret it, that the psychological world as well as the university world has gone far toward recognizing not only the problem, but the products of the ESP research—far, at least, as measured by the state of affairs in the days of James and McDougall.

But if parapsychology, as represented by the research in extrasensory perception, has clearly achieved a place in the university world and in the psychological laboratory, it has supplemented rather than supplanted the work of the societies of psychical research. And the work carried out between the societies and university staff members both in England and America gives ample assurance that cordial and profitable cooperation may be counted on between these different institutions of research. Such cooperation is nowhere better exemplified than in the joint endeavors of the American Society for Psychical Research and the parapsychological work going on under Dr. Murphy at the College of the City of New York, aided, from a third source, by the Hodgson Fund for Psychical Research administered by Harvard University.

There are other objective signs of the dawn of greater opportunity for this field of inquiry on the university campus. For example, readers who are familiar with the many contributions of Dr. C. E. Stuart, formerly of the Parapsychology Laboratory of Duke, will regard his appointment to the Fellowship in Psychical Research at Stanford University as a progressive event. Likewise, the fact that Harvard Summer School this year offered a lecture course under Dr. Gardner Murphy on the subject of psychical research as a part of the William James Centenary is good news for the field. The granting of the Ph.D. degree for parapsychology theses by London and Duke universities in recent years and the direction of research in this field by committees of distinguished staff members at Cambridge and Harvard universities contribute further to the evidence of growing recognition.

These events will appear as small achievements when compared

to the developments in more favored fields, but when viewed in the light of the long struggle of the pioneers who have gone before us, they give ground for real encouragement and optimism for the future, at least in so far as the ESP research is concerned. And there will be no great difficulty, we think, with other problems in parapsychology if they can be approached and investigated with the same standards of scientific caution as has now become established in the ESP research.

AN EXPERIMENT IN PRECOGNITION USING DICE

By J. L. Woodruff and J. B. Rhine
Parapsychology Laboratory, Duke University

Abstract: This paper describes an experiment made to find out whether the fall of a die can be predicted with results that could not reasonably be ascribed to chance.

In one of the two main series, the prediction of the face to come up when the die was cast was made just before each throw of the die. In the other series the predictions were all made for 24 throws of the die before the throwing was begun. The main difference between the two series lies in the fact that the 1-Throw Series received a throw-by-throw check-up. This condition is probably important because it considerably enlivened the test. The results of this series were highly significant for one of the two subjects who participated. Where the chance average is 4 hits per run of 24 trials, the subject's average score was above 5 hits per run, a result which gave a critical ratio of 4.78, well above the 2.5 standard of significance. This figure, representing the performance of a single subject in only one of the two series, is enough to establish the extra-chance character of the work as a whole. Even when the probability is found for combined results of the whole research, P is still approximately only .0001. In other words, the odds are about 10,000 to 1 that such a result would not occur by chance alone.

A number of alternative explanations are considered in the light of the results and are found to be inapplicable. For example, there is the possibility that the die was thrown with some trick or skill that made it fall as predicted. But the fact that the best results were given when the throwing was done by an electrically driven rotating cage obviates that possibility. Likewise the hypothesis that the results may have been due to faulty dice was considered and statistical treatment submitted to show that this hypothesis will not fit the facts of the case.

A striking decline of success in scoring occurred from the first half of the run to the second, and this effect holds for both subjects and both series (i.e., in the four subdivisions). The ratio of deviations from the first to the second half of each run of the main series as a whole is 68 to 6.

It is concluded that the precognition hypothesis is the one most applicable to these results and that this hypothesis has been found to be demonstrable by the die-prediction technique—as successfully, it may be added, as by any of the other methods reported. The present work not only adds to the list of confirmations, but further increases, too, the number of experimenters contributing to the case for the occurrence of precognitive ESP.

OF THE TEN reported investigations of ESP which bear evidence supporting the precognition hypothesis, six have been based upon some modification of the ESP card tests.[1] That card tests should have been used predominantly is explainable since the experimental investigation of precognition (Pcg) arose out of the more elementary investigations of ESP, for of course the Pcg hypothesis is nothing more nor less than ESP of future events.

The other four reports deal with different methods of testing the hypothesis. That by Saltmarsh was based entirely upon the analysis of spontaneous experiences of previsionary character (14); one by Carington arose incidentally out of tests for telepathy using drawings, the subjects sometimes getting ahead of the selection of the drawing and reproducing it in advance (2). Tyrrell's work utilized the technique of identifying the selection made by an electric machine in advance of the selection (16). And only one, decidedly the least impressive of all (also by Carington), utilized the throwing of dice with the subjects predicting which faces would appear when the dice were cast (1).

The investigation we are about to report employed the method last mentioned, that of predicting the face of a die that would come up when the die was thrown in random fashion. Our investigation of Pcg with the use of dice was embedded in a series of investigations of

[1] All of these are from this laboratory: Rhine (11, 12, 13); Hutchinson (7); Stuart (15); Humphrey and Rhine (6).

statistical and psychological character (other than ESP) based on dice-throwing procedures that were begun in February, 1934. These investigations had been under way well over two years when the prediction tests reported here were undertaken. There was doubtless some stimulation as a result of the earlier dice-prediction report (which came out in June, 1935) but the fact that this laboratory was actively concerned at the time with the investigation of Pcg and with the studies of dice throwing that had been going on for some time made the convergence of the two lines natural in the present experiment.

The experiments to be reported were carried out well ahead of most of the reports on Pcg that have already been published, and this raises the question as to why this report has been withheld up to this time. The answer is that the major research with which these experiments have been associated has not been regarded as ready for publication hitherto and it has seemed advisable to keep the logical association of the Pcg work and that of the larger research project relatively close and apparent in the order of publication as well. But since the time is approaching in which the whole project will be made available in a series of reports, there is no longer need to withhold the Pcg phase of this investigation. However, in order to avoid endless confusion of issues, it is necessary to confine the present report entirely to the Pcg hypothesis. The task of placing it in its larger setting may be left to a subsequent review when the total research picture is available.

The Experiment

The work here reported was done in October of 1936, from the 9th to the 26th, inclusive. The tests were conducted under the immediate supervision of JLW who was then Graduate Research Assistant in the Parapsychology Laboratory. JLW participated also as a subject in the tests, sharing this function with A. J. Linzmayer (AJL), formerly on the secretarial staff of this laboratory. Both were present during all of the tests which form the main body of this report. A small section of data obtained with JLW alone, a short series by AJL with another observer than JLW, and a few runs made by two other subjects will be mentioned as a miscellaneous section to complete the report, but will not be given full consideration.

The main body of tests, thus, comes from an experimental situation in which JLW and AJL alternated as subject in predicting the face of the die that would come up on the throw. It will be seen below that in most of the work the other participant took over the task of throwing the die, and throughout the main experiment he was always the observer.

The die used was a common white cube about $^{11}/_{16}$ of an inch on the edge made of plastic material marked with black spots. It was thrown in two ways, each indicated in the report of the results presented below: The first was the hand-thrown method in which the die was thrown from a round cardboard cup of about a quart capacity on to a special table, 3 feet by 5 feet, covered with soft padding on which the die rolled and bounced easily. The table had side walls, 6 to 12 inches high, also padded. It was required that the die either be shaken in the cup or bounced against the side walls of the table. The second way of throwing the die consisted of putting it in an electrically rotated wire-mesh cage (4 by 4 by 25 inches) and at each half-revolution observing the face landing upward on the wooden ends of the cage. The cage was equipped with various randomizing devices (mostly cross wires or bands) to render the course of the die more erratic as it rolled from one end to the other. The rate of the machine was about 5 seconds per half-revolution.

Two general divisions of the data will be made according to the principal conditions of the experiment. In the division called the 1-Throw Series, the subject predicted the fall of the die just prior to each throw. That is, a die face would be named (say, "three") and recorded before the die would be cast or the machine reach the point where the die fell. Then the die would be thrown, or would fall, it would be read, and the number called aloud and recorded. Both subject and observer could then watch the rate of success maintained as the test progressed, a practice which always adds interest.

In the other division, the 24-Throw Series, a complete column on the record sheet (i.e., a run of 24 trials) was filled out with the subject's predictions before the die would be thrown for the first trial. Then the die would be thrown 24 times in succession and the readings recorded in the adjoining column. Again the subject and observer could both watch the degree of success, *but only as the throwing progressed*, not as in the 1-Throw Series *during* the prediction. (In

one small section of this series—unimportant, as will be seen, to any conclusion drawn in the paper—the throw record for one subject, JLW, was not taken in full; only the hits were checked as the thrown face coincided with the prediction record.)

The recording was done at a period before the laboratory instituted its practice of independent records. It was done by the subject doing the predicting. He simply wrote the face-number predictions (from one to six) in the proper spaces on the record sheet, and as the die was thrown, he called the uppermost face aloud and recorded it in the adjoining column. His record sheet was, throughout, in good view of the other subject who played the role of observer, as well as threw the die. A friendly, though lively, competition was carried on between the two involving small wagers on each run. JLW recorded in ink (and his records are the only significant ones) while AJL used pencil. The subject, witnessed by the observer, made the check-up after each run and summarized the scores on the record sheet. A summary and report were made by JLW to JBR immediately after the research was completed.

There were 304 runs (each of 24 single die predictions) made in the main experiment; 126 in the 1-Throw Series and 178 in the 24-Throw; 131 by JLW and 173 by AJL. The Miscellaneous Section contains 78 runs of the 1-throw and 51 of the 24-throw type. (A block of 44 runs reported in the summary have not yet been located but, as will be seen below, there is no difference in the conclusions with these omitted. Their inclusion would raise the CR.)

The records of all the runs listed in the summary which were accounted for were rechecked and the analyses made by Miss Betty M. Humphrey, Graduate Assistant of the Laboratory, during the summer of 1942, and they have been completely rechecked for computational errors.

The various subdivisions of the results are indicated in Table 1 below, and the size of each is shown by the number of runs. Each run represents 24 predictions, or one column on the record sheet.

Results

In the presentation of the results, it will be of some advantage to review the totals of the data, since this affords a perspective for consideration of the special sections which are of greatest interest. Alto-

Table 1

RESULTS OF DIE-PREDICTION TESTS

A. MAIN EXPERIMENT

Thrown by	1-Throw Series: Subject Predicting JLW Runs	JLW Dev.	AJL Runs	AJL Dev.	24-Throw Series: Subject Predicting JLW Runs	JLW Dev.	AJL Runs	AJL Dev.
AJL	40	+38	..	....	..	....	..	....
JLW	10	+10	19	+1	28	+10	81	+16
Cage	20	+25	37	−4	33	−10	36	−12
Subject Total	70 CR = 4.78	+73	56	−3	61	0	117	+ 4

	Runs	Dev.	CR	Runs	Dev.	CR
Series Total	126	+70	3.42	178	+4	Insig.

B. MISCELLANEOUS TESTS THROWN BY VARIOUS METHODS

	1-Throw Series: Subject Predicting JLW Runs	JLW Dev.	AJL Runs	AJL Dev.	Others Runs	Others Dev.	24-Throw Series: Subject Predicting JLW Runs	JLW Dev.	AJL Runs	AJL Dev.	Others Runs	Others Dev.
Subject Total	27	+26	23	−3	28	+9	20	+1	21	+1	10	−14

	Runs	Dev.	CR	Runs	Dev.	CR
Series Total	78	+32	1.99	51	−12	.92

C. TOTALS BY SUBJECTS AND BY SERIES

	1-Throw Series: JLW Runs	JLW Dev.	AJL Runs	AJL Dev.	Others Runs	Others Dev.	24-Throw Series: JLW Runs	JLW Dev.	AJL Runs	AJL Dev.	Others Runs	Others Dev.
Subject Totals A and B	97 CR = 5.51	+99	79	−6	28	+9	81	+1	138	+5	10	−14

	Runs	Dev.	CR	Runs	Dev.	CR
Series Totals A and B	204	+102	3.91	229	−8	Insig.

SERIES GRAND TOTAL....	Runs	Deviations	SD	CR
	433	+94	±37.98	2.47

gether, there were 433 runs of 24 trials each made in the die-prediction tests, 204 of them being by the 1-throw and 229 of them by the 24-throw procedure. These 433 runs have a chance expectation of 4 hits per run or 1,732. A total of 1,826 correct predictions were made which is 94 above expectation. The standard deviation is ± 37.98 and the CR is therefore 2.47. Thus the work as a whole almost exactly meets the 2.5 criterion of significance.[2] It has a probability of about 150 to 1 that such a deviation would not occur by chance alone.

But such a general pooling of the data is not essential to their analysis and does in fact obscure the main points of interest, even the very striking extra-chance effect which was obtained. This effect, as may be seen in Table 1, appears entirely in the work of one subject, JLW, and only in the 1-Throw Series. While the 229 runs of the 24-Throw Series give a deviation of —8 (8 points below expectation), the 1-Throw Series has 204 runs with 102 hits above expectation. This gives a CR of 3.91 and is therefore highly significant.

There are included in these totals, however, a block of runs of miscellaneous tests (shown in Section B of the table) which are not regarded as an integral part of the main experiment (Section A). Their inclusion does not alter the conclusions in any way—indeed, they tend strongly to confirm the results of the main division—but the conditions were not in all cases as well controlled as those described above for the principal research. They are mentioned merely to place on record all tests conducted at the time as part of the die-prediction experiment and are included in the table for their comparison value. They total 78 runs for the 1-Throw Series, with a deviation of + 32, and 51 runs for the 24-Throw Series, with — 12 as a deviation.

Omitting consideration of the Miscellaneous Section and turning attention to the data summarized in Section A of the table, we present the results of the main experiment in the two series. The 1-Throw Series again is the significant one, giving a CR of 3.42 for the 126 runs with a deviation of + 70. The 24-Throw Series is insignificant, having only a deviation of + 4 in 178 runs.

Subdivided still further, on the basis of subjects, the 126 runs consist of 70 runs by JLW and 56 by AJL. JLW obtained a devia-

[2] The 44 runs for which the original records have not been located as yet would bring the CR above this criterion, according to the original summary of the results.

tion of + 73, *averaging above 5 hits per run for 70 runs*, as against a chance expectation of 4! The SD is ± 15.27 and the CR the quite outstanding figure of 4.78.

(In passing, it should be pointed out that JLW's work in the Miscellaneous Section was at the same general average of scoring as his Main Section tests, and this is true both for the 1-Throw and the 24-Throw [see Table 1, Section B]. Taken together JLW's work in the two sections totals 97 runs at the average of 5 hits per run, giving 99 hits above expectation and a CR of 5.51. AJL's work yields approximately the same average in the two sections; and again in both series [1-Throw and 24-Throw].)

Taken in any of the several groupings mentioned, JLW's results are highly significant and there is no required correction that would alter this situation materially. Even if his results for both sections and series are pooled there is still a CR of 4.11 from the 178 runs with a + 100 as deviation.

The chi-square method may be used to combine the CR's of the different subdivisions, the work of JLW and that of AJL. This method preserves the outstanding character of the JLW work of the 1-Throw Series that is lost to a great degree in pooling the deviations for the entire work. Taking the four subdivisions of the main experiment as shown in Section A of Table 1, (i.e., work of both subjects for both series) the sum of the chi-squares (CR's squared) is 22.94. With 4 degrees of freedom, the probability of this is less than .0002. Or if we include the two subdivisions of the Miscellaneous Section, the P-Value is even a little less, deriving from a total chi-square of 27.75 with 6 degrees of freedom; P = .0001.

Considered thus from any proper statistical point of view, the performance of JLW, mainly in the 1-Throw Series, is beyond doubt an extra-chance occurrence, even when duly corrected for its selection from the total work on die prediction.

There are, however, other questions to consider before a conclusion would be in order concerning the Pcg hypothesis. Chief among these are "optional stopping," possible defects in the die combined with preferential prediction by the subject, clerical errors, and improper selection of data. After a discussion of these questions, we will then consider the interpretation of the results as "extra-normal,"

and attempt to show how far the Pcg hypothesis is a more fitting one than the Pk effect.

Counter-Hypotheses

A proper question to ask at the outset, because it is related closely to the chance hypothesis just dismissed, is: Do the data represent a proper sample for statistical treatment? The following general information is relevant to this question: The whole of the experimental results obtained under the conditions is submitted as a sample without selection except as described and accounted for. Such subdivision as is made is accounted for in the foregoing text. The omission of a miscellaneous section of data from each of the main series is explained, and the alternative of inclusion is observed to make no difference in the statistical conclusion reached. Hence, this report concerns a unit of research data complete in itself.

There remains, insofar as sampling is concerned, only the question of optional stopping, one that might be raised in some form in nearly all statistical work, but which is of real concern mainly in research such as this where every remotely conceivable alternative to the hypothesis under test is to be scrutinized with utter exhaustiveness. The question itself was not expressly "in the air" at the time, and provisions made later for avoiding this issue had not then been taken; namely, the setting in advance of a fixed number of runs as the stopping-point. But from the conditions of the experiment and the results, we may still satisfactorily answer the question: Did optional stopping operate to produce the significant results reported? To facilitate discussion, let us formulate the issue more sharply: Did the die-prediction work merely begin and continue *as a chance series* until a significant positive deviation was reached and stop arbitrarily at that favorable point? The answer is manifest in the records. First, the work as a whole, when pooled, gives a CR that is just barely significant; but this was not known at the time, and the number of omission errors would have held the CR below 2.5 had the computation been made then. Second, the first eight days' work, from October 9 to 16, all of it done by JLW, gave a positive deviation of 48 in 50 runs (CR of 3.72). This is ample evidence that from the start the experiment was not of chance character. This fact effectively bars the application of the optional stopping hypothesis. Third, not only was

JLW successful from the beginning, he held up his score average to the end (CR of his total = 5.51). Obviously, JLW's work, on which the whole case against chance rests, was not at any point a chance series. This dismisses the question of optional stopping.

The research interest had centered rather in the individual performance of JLW and in the contrast of his work with AJL, as well as in the comparison of 1-throw with 24-throw tests. And finally, the general effect of this work against the larger background of the research program of the laboratory which has yet to come into publication was of importance. Stopping was actually determined when these larger objectives seemed to be reasonably satisfied and when a requisite degree of individual satiation was reached by the subjects themselves.

But even though the work constitutes a proper sample in itself, may not the significant result be due to a concurrence of a peculiar distribution of die faces and of similarly distributed predictions by JLW? Pointing the question still further, might not JLW have followed, unconsciously perhaps, an assumed and distorted distribution of the faces thrown? Suppose a given face is noted to come up more frequently than the others; then simply by predicting this one more frequently, an obvious increase in successes is likely to result, roughly in some proportion to the excesses of the throws and the calls of the face in question.

This question of distribution similarities can likewise be answered from the data themselves. From JLW's records, which are the point of focus of interest, the frequency both of the predictions of the different faces and of the actual die-throws has been determined. Then for each face, the empirical expectation of hits is found as follows: Number of predictions for a given face over total number of predictions for all faces, times the number of throws giving that particular face. This empirical expectation is then subtracted from the actual number of hits made in that face. This is done for each of the six faces, and the six deviations pooled. For JLW's 70 runs of 1-throw tests, the empirically found deviation is + 67. This is sufficiently close to the deviation of + 73 from the theoretical mean to enable the present issue to be dismissed. It is also 3.87 times the empirical SD of ± 17.33. However biased JLW was in his predictions, and however crooked the die may have been, it requires more than mere

peculiarity of the mass distribution of face-frequencies to account for the highly significant effect produced. (Further details of the computation referred to above may be seen in Table 2.)

Table 2

DEVIATION COMPUTED FROM EMPIRICAL EXPECTATION IN JLW'S 1-THROW RESULTS

Face	No. Pred.	No. Thrown	Empirical Expect.	Hits	Dev.	$\frac{\text{Dev.}}{np}$
6's	425	305	77.16	101	+23.84	.309
5's	320	296	56.38	71	+14.62	.259
4's	283	273	45.99	52	+ 6.01	.130
3's	242	287	41.34	48	+ 6.66	.161
2's	274	282	45.99	57	+11.01	.239
1's	136	237	19.19	24	+ 4.81	.251

Total deviation = 67 ± 17.33 CR = 3.87

As a third counter-hypothesis, there is still the hypothetical possibility that the die could have been so "manipulated" as to give other than chance results, that it might have been skilfully controlled so as to hit the predictions to a better-than-chance degree. This hypothesis need not suppose conscious manipulation nor any deliberate intent to deceive. But this possibility is eliminated effectively by the methods of throwing the die. Even if it be supposed that the cup throwing could be effectively controlled, and it is difficult to see how it could be, the rotating cage can hardly be supposed to be subject to this hypothesis. And it was with the cage that JLW obtained his most spectacular deviation: + 25 on 20 runs.

Something might be said about the recording and the clerical procedures of the research. A few years ago a great deal of critical discussion was started over the recording error hypothesis as a possible explanation of the ESP results—a criticism which, as was shown at the ESP Symposium before the American Psychological Association in 1938 (3), was based on an astonishingly meager foundation. But it is not a wholly pointless issue, and the average experimenter has since preferred to take precautions far exceeding general research standards in other fields rather than allow any question to be raised as to recording error possibilities. The present work was done before this strict methodological innovation was adopted. Nevertheless, it may be possible to clear up whatever questions may arise on this point.

First, the conditions were moderately secure against such errors.

As stated above, there were two persons present in all the work of the main experiment, both of them experienced in the work being done, and both on the staff of the laboratory (one in a research position, the other, secretarial). They were lightly competing with each other for small stakes, to be paid from their own pockets; this presumably provided a certain amount of increased motivation to observe the recording and check the scoring. The recording was always done in full view of both the subject who predicted (and recorded) and the one who threw the dice (or ran the machine).

No direct measure of accuracy is possible in the case of JLW's recording, which is the crucial part of the research. But we have the following indirect reflection on the question: In AJL's recording there were 31 runs in which there were double records made in order to make each throw of the die serve for two predictions. These records can be checked against each other for inconsistencies, as a measure of accuracy. There were 744 double entries, and of these, two were in disagreement. This is a rate of error of approximately .003. We cannot be certain whether this is an accurate measure of AJL's total recording errors, but it may be fairly close. Now it is possible to compare JLW and AJL on the other types of clerical errors which can be readily estimated from the recheck which all the data have been given in the preparation of this report. And from these other, but related, data on accuracy, it appears that JLW is the more accurate of the two, as his greater academic experience and psychological training would naturally lead us to expect. In fact, he made only about one third as many errors as AJL. It seems safe to infer, then, that there is very small danger of recording errors made by JLW accounting for his consistently high scoring in the 1-Throw Series. In this series, on which the significance of the experiment depends, each man made only one error; JLW, an error in checking, and AJL, one in counting the hits. Both were omissions and lowered the scores, as is a general rule in the studies of clerical errors reported (5).

Most of the errors were made in the 24-Throw Series, with a few also in the Miscellaneous Section, and these were mostly later on in the series, and as stated, mostly by AJL. JLW made 5 checking errors, 4 of them omissions, while AJL made 14, 13 of them omissions. AJL made 1 counting error. In the Miscellaneous Section, AJL

made 6 omissions in checking and JLW, 1. One other subject made 1. This point is clear: these errors tend to be on the "safe side," lowering rather than raising the score.

The relatively higher error-rate in the 24-Throw Series and Miscellaneous Section, as compared to the 1-Throw Series, may be reasonably ascribed to the fact that much less interest and importance were attached to those tests. It was the 1-Throw Series that was giving a deviation and hence was relatively more important; the other series soon became just a control, with chance scores as the rule. Lowered attention would favor more errors. Besides, the recorders knew there would be a later recheck of all the records; the immediate check was for the demands of the moment, and these, except for the 1-Throw Series, were of little interest and were naturally done with less time and attention. The 1-Throw Series, then, affords a better basis for judgment of the accuracy of the recorders and by this, they both measure up well; in fact, equally well.

Returning now for a final word on the recording error hypothesis, we will consider briefly the question of "motivated" error. That we do so frankly is due to the fact that we are aware of the tendency of a hard-pressed critic to ask, as a last resort, whether the recorder's bias may not lead to errors (conscious or otherwise) that favor his hypotheses. JLW would be the center of interest, again, on this issue. But we still have the question as to why he would have made such hypothetical errors only on the 1-Throw and not on the 24-Throw Series. We know of no bias that would have favored the one against the other. In fact, it would have been a greater feat, a more interesting result, to do well in the 24-throw test; and hence the 24-throw test should have "spurred" the "erring motivation" into activity more than the 1-throw. We must recall, too, that JLW was being observed continually in the Main Series and that there was a degree of competition involved. Finally, the "motivational theory" of error is badly shaken by the nature of the errors we find on the recheck: First, the fewest recheck errors were found in the most significant sections; second, of the total of 31 errors made, only 3 were of the score-raising type and 28 lowered the score. (In the 3 score-raising errors was one that consisted of an omission of a run, which raised the score by four points, making up 6 points as against 28 points omitted by error.)

Loss of records is likewise a possible danger in such research, especially since possible shift of significance may be effected in this way. But a summary of the work done was made at the time and this allows a complete check on all the tests recorded. There is, in fact, a block of 44 runs that has not been located as yet. (Six years have elapsed since the records were taken and changes in the record filing system have been made.) But of these, 23 runs were made by JLW in the 1-Throw Series, and the summary sheet shows a deviation of +17 for these. (However, since the 70 runs found give a CR of 4.78, this is of no great consequence.) The misplacement of these records, then, serves to weaken rather than strengthen the statistical case. The other 21 misplaced runs were done by AJL (1-Throw) and show a deviation of −10. The conclusions, then, would be the same whether these records were eventually located or not. (Their inclusion in the grand totals would raise the CR of 2.47 to 2.53.)

Readers of earlier papers on Pcg from this laboratory will recall that in interpreting the experimental results as bearing on the Pcg hypothesis, the logical possibility was considered that some principle of psychophysical causation operated to *produce* the events predicted. This was done in spite of the lack of any scientific recognition of this hypothesis because, in entertaining the precognition hypothesis itself, we began with little more than just such a logical possibility. We are accustomed in parapsychology to dealing expressly with a wide range of counter-hypotheses which ordinarily are not considered. (Involuntary whispering, collusion, errors, lost records, selection, and many others. How many of these have been considered seriously in other fields?) This hypothesis of a "psychokinetic" influence over events was excluded, presumably by certain experimental procedures aimed at keeping strict mechanical control over the events (the shuffling of cards).

In the dice-prediction work, no such preventive measures were taken, but it is believed that the counter-hypothesis can be rejected on the basis of rational consideration of the results. For example, JLW scored well in all three of the conditions under which the die was thrown—cage throwing, AJL throwing, and himself throwing. If anything like the hypothetical psychokinetic principle were active, surely in the change from JLW himself to AJL as thrower, there should be expected a big change in scoring. But there is not much

difference; the average score changed only from 4.90 to 4.95. And furthermore, changing over from a *person* throwing the die to a *machine* throwing would surely put the hypothetical effect under some new influences that would show up in greatly altered averages. But again, it changed only from 4.95 to 5.25 which is a relatively modest change. (The runs and deviations for these three conditions may be seen in Table 1.) On the other hand, the process of prediction itself presumably is not much concerned with these changes of type of throwing; hence the stability of JLW's score averages argues for the precognition hypothesis and against Pk as applying here.

Moreover, if JLW exercised any of the supposed psychokinetic factors in the 10 runs that gave + 10 in his prediction tests, how can we account for his giving only + 1 on 19 runs in which he threw the die for the prediction tests of AJL? If we were to assume this hypothetical principle as determinative, then it is necessary to account for this and similar inconsistencies. Why, for instance, should AJL, on such an hypothesis, give JLW's predictions on 40 runs a + 38 deviation, yet exert no "influence" on his own? Only if we suppose that JLW obtained his consistently high scores by Pcg can we see how the changes in the throwing, above all the transition to the machine, need not be expected to affect the scoring.

Finally, JLW failed to score well in the 24-Throw Series. The throwing aspect was not much different, and in fact *the throw distribution (of faces) is about the same for the two series.* But the prediction conditions are much more distinct. For that matter, to make a long run of *predictions* (24 at a run) without waiting for a check-up is very different from having each one checked before going on to the next; whereas in the *throwing*, the check-up is obvious on each throw in both series and in this they are alike. This argues that it is probably the much greater difference in the prediction aspect of the tests that produced the great difference in scoring; that it is the precognition process that is responsible instead of any influence over the throwing.

In the light of these features of the evidence, and in the absence of any positive evidence in the results for the counter-hypothesis, it is probable that all readers will concede the case against a direct psychokinetic effect on the dice and allow the conclusion, instead, of a

precognitive relation between the subject's predictions and the falling of the dice.

One of the most reassuring features of the experimental results is a decline effect which the analysis revealed—a drop in scoring level from the first half of the run to the second. Each run consists of 24 trials, and when the hits of the first 12 trials of all the runs of the 1-Throw and 24-Throw Series combined are pooled and compared with the same figures for the last 12 trials of the run, the results are:

First Half		*Second Half*	
Hits	676	Hits	614
Dev.	+68	Dev.	+6

The difference is not significant (CR of Diff. = 1.95), but the ratio of deviations, 68 to 6, is itself strongly suggestive.

This drop effect prevails in the subdivision as well: The 1-Throw Series drops from 303 hits on the first to 271 hits on the second half (+51 to +19); and the 24-Throw from 373 to 343 (+17 to −13). Even the subjects individually show the same relation as follows:

1-Throw Series					24-Throw Series				
	JLW		AJL			JLW		AJL	
	1st half	*2nd half*	*1st half*	*2nd half*		*1st half*	*2nd half*	*1st half*	*2nd half*
Hits	180	173	123	98	Hits	133	111	240	232
Dev.	+40	+33	+11	−14	Dev.	+11	−11	+6	−2

Thus, the effect appears not to be random. The reason this drop is regarded as "reassuring" is that a general study of declines within the run and other test units is being made (at the time of writing, in September, 1942), and the decline effect appears to be sufficiently general to be designated as almost characteristic of the presence of the ESP effect. It will be recalled that it occurred in much of the earlier ESP work. Richet (10) noted it in his long runs, Estabrooks (4) found a great drop in the second half of his runs, Jephson (8) saw a decline in a 5-trial run, and Rhine reported two cases of such declines in his subjects AJL and HLF. (See graph p. 141, *Extra-Sensory Perception.*)[3] Martin and Stribic (9) found a similar decline and other instances are now being discovered as the older data which are comparable are re-examined with this in mind. The significance of

[3] Rhine, J. B. Extra-Sensory Perception. Bruce Humphries: Boston, 1934.

this decline effect in the present report is largely its indication of something at work beyond the range of the common counter-hypotheses. At least none of the common explanations will account for such decline effects. There will be much more to say in later papers regarding these declines, their frequency, and the conditions that seem to produce them.

Discussion of Further Points of Interest

Why did not AJL score as high as JLW? Why did not the 24-throw test work as well as the 1-throw? Why do we find the peculiar distribution of faces shown in the results of JLW's tests as they appear in Table 2? These are questions that should be recognized and raised even if to give them only the most tentative answers for the present.

The examination of AJL's prediction records suggests an adequate and possible explanation for his failure. He was much given to a patterning of his predictions, to repeating these patterns frequently, and apparently to following a relatively closed chain of sequences of faces. This would work against precognitive ESP by substituting a response obtained by association or habit instead of ESP. Even from the statistical viewpoint, he handicapped himself since the distribution of his predictions (see Table 3) shows that he greatly neglected fives and sixes and overcalled twos and threes, and since the die-throws were not distributed in the same peculiar way. These practices may not have been the only factors, and scoring above chance would be possible in spite of them; but they are very probably contributory to the chance results obtained by AJL.

Table 3

Distribution of AJL's Predictions in Terms of Deviation from Mean Chance Expectation

Series	Face						
	1	2	3	4	5	6	M.C.E.
1-Throw	− 16	+ 85	+55	+ 18	− 82	− 60	224
24-Throw	+116	+160	+ 9	+111	−268	−128	468
Total	+100	+245	+64	+129	−350	−188	692

The failure on the 24-Throw Series too may be explained in a tentative way. It was naturally less interesting because of the lag in checking up on success, and probably this fact is enough to account for its failure to produce high scoring. Especially was the adverse contrast with the 1-Throw likely to create a preference for the latter. Being less interesting is generally regarded as a great handicap in an ESP method. Also, delay is frustrating to the subject, especially if he has also to go ahead and write down a lengthy run of calls instead of checking up as he goes. The point is, that after the subject has already been accustomed to the more gamelike test of the 1-Throw Series, the 24-Throw would be slow and worklike in comparison. Taken alone, it might have been different.

The most plausible, certainly the easiest, answer to the third question, Why the peculiar distribution of faces thrown in JLW's work? is that the die is imperfect. Were it not for the fact that it did not always give the same type of distribution, the imperfection hypothesis would appear very likely. But the distribution from the cage throwing rather argues against the imperfection, though not with finality. That is, while JLW's 70 runs gave 43 fewer 1's than the 280 expected, and this with a SD of $\pm$ 15.27 gives a CR of 2.82, the total of 93 runs thrown by the cage (for which die records are complete) give no significant deviation on any face. It is, however, noteworthy that the face most outstanding in deviation *is* the one-face, and it *again* has negative deviation. This time it is -36 ± 17.6, with the CR 2.05. The question involves features not easily introduced on the basis of the data and the analyses available in the present paper. Accordingly, whether the die is imperfect (disfavoring the 1-face) or whether it was slightly influenced in the unrecognized manner called Pk (again disfavoring the 1-face by reason of a "preference" for other faces) cannot be settled here. Obviously, most readers will find it easy to prefer the hypothesis of imperfections in the die. It is sufficient here, however, to recall that even when the actual number of faces coming up was used to compute expectation, there was still the significantly large deviation of $+67$. And whatever the cause of the distribution, this would be independent of any inequality of faces thrown. The case for precognition, therefore, stands clear of any question of the peculiarities of frequency distribution of the die-throws or, for that matter, even the predictions.

Admittedly there is suggested by the unusual distribution of faces thrown the possibility of some factor influencing the fall of the dice, even of the much-rejected Pk hypothesis. But we can go no further into that question through the analysis of these results than to say that it is raised as a question for study. If these tests provoke this "old" parapsychological question, others better adapted for the purpose than the present work may be found to answer it. That is as far as this report can properly go into the Pk hypothesis.

So far as the case of Pcg is concerned, the present report really adds little beyond another confirmatory research. As stated at the beginning, the work was a relatively early effort, and it took up the question at a primitive stage, as things now seem. This is not to minimize its value; at this point of the Pcg research even mere repetition is worth a great deal, and in addition, this paper introduces a new investigator (JLW) into the literature on that topic.

But a few points of value have been contributed to our knowledge of Pcg, especially to the conditions of its occurrence. The decline effect, the drop in deviation from the first to the second half of the run, is of special interest in view of its comparative value for other work and this interest may very well become much greater in the future. The success of JLW, no matter whether the dice were thrown by himself, by AJL, or by the machine, emphasizes the independence of the Pcg effect of what is happening to the objects involved. And finally, the complete failure of JLW on the 24-throw test makes clear how the slenderest margin of difference in the test procedure may completely and consistently bar the subject from success.

Conclusion

The significant results given by JLW in the 1-Throw Series lends an extra-chance character to the investigation as a whole. The probability for the total results is about .0001.

A number of counter-explanations are considered and rejected, leaving only the precognition hypothesis as acceptable. This is, then, another research unit to add to the growing support of the case for precognition.

A marked decline of success from the first to the second half of the run is found in the total results, a decline that is found also in both series and with both subjects. While not significant, this great

drop in scoring is of particular interest in view of similar effects found in other work, published and unpublished, which are scheduled to be reported in due time.

References

1. Carington, W. Preliminary experiments in precognitive guessing. *Jour. S. P. R.*, 1935, **29**, 86-104.
2. ———. Experiments on the paranormal cognition of drawings. *Proc. S. P. R.*, 1940, **46**, 34-151.
3. ESP Symposium at the A. P. A. *J. Parapsychol.*, 1938, **2**, 247-272.
4. Estabrooks, G. H. A contribution to experimental telepathy. *Bulletin Boston S. P. R.*, 1927, **5**, 1-30.
5. Greenwood, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, **2**, 138-146.
6. Humphrey, B. M., and Rhine, J. B. A confirmatory study of salience in precognition tests. *J. Parapsychol.*, 1942, **6**, 190-219.
7. Hutchinson, L. Variations of time intervals in pre-shuffle card-calling tests. *J. Parapsychol.*, 1940, **4**, 249-270.
8. Jephson, Ina. Evidence for clairvoyance in card-guessing. *Proc. S. P. R.*, 1929, **38**, 223-268.
9. Martin, D. R., and Stribic, F. P. Studies in extra-sensory perception: III. A review of all University of Colorado experiments. *J. Parapsychol.*, 1940, **4**, 159-248.
10. Richet, Charles. La suggestion mentale et le calcul des probabilités. *Revue Philosophique*, December, 1884.
11. Rhine, J. B. Experiments bearing on the precognition hypothesis. *J. Parapsychol.*, 1938, **2**, 38-54.
12. ———. Experiments bearing upon the precognition hypothesis. III. Mechanically selected cards. *J. Parapsychol.*, 1941, **5**, 1-57.
13. ———. Evidence of precognition in the covariation of salience ratios. *J. Parapsychol.*, 1942, **6**, 111-143.
14. Saltmarsh, H. F. Report on cases of apparent precognition. *Proc. S. P. R.*, 1934, **42**, 49-103.
15. Stuart, C. E. An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *J. Parapsychol.*, 1941, **5**, 99-137.
16. Tyrrell, G. N. M. Further research in extra-sensory perception. *Proc. S. P. R.*, 1936, **44**, 99-168.

SOME OF THE CONVERGENCES WHICH INITIATED PARAPSYCHOLOGY

By GERALD HEARD

In the June issue of this Journal, the question was raised as to why parapsychology (or psychical research) became a matter for serious scientific research only in the last quarter of the nineteenth century. This is no idle inquiry. On the contrary, it is one of those lines of research, or para-experimental investigation, which are essential if, in any of the sciences, but especially in a new one such as parapsychology, we are to understand the conditions under which research can most favorably be conducted. These conditions are little known, less studied, but, many inquirers would now agree, more important than most onlookers would imagine. In his *Great Society*, Professor Graham Wallas, a generation ago, drew attention to the unsuspected influence which changes in social conveniences and comfort made, in the early modern age, to the quality, range, and variety of subjects into which people began then to inquire. A decade ago, Professor Hocart, in his *Progress of Man*, demonstrated, further, how specific observations of certain phenomena can be made over great areas of time, but they will lead to no true advance in man's thought unless and until there is a change in the intellectual climate which permits these facts to be seen as part of an integral order. So, though Hero of Alexandria invents a turbine in the first century A.D., this and other pressure-engines remain sterile fancies until Boyle's theory of gases makes men's minds accustomed to thinking of the play of orderly forces. This factor of the intellectual climate the June editorial raised for consideration. As it is of such importance and since in that survey not all the elements were mentioned, permission has been granted to develop the theme a little further.

And, indeed, the more that last quarter of the nineteenth century is studied, the more surprising it is to see what a number of "convergences" then contributed, from so many alien interests, to make

possible research into parapsychology. There was a rapid "fitting of the environment" so that such studies could be undertaken. The difficulties to be overcome were immense. By the middle of the century, the sciences had arrived at a hardened assurance that in materialism, or at least mechanism, was to be found the invariable law explaining all phenomena. Wöhler's synthesis of urea had dealt a severe blow to old-fashioned vitalism. *The Origin of Species* had seemed to explain life in terms of a law which was only mechanism applied to reproduction. So dogmatic had scientists become that the curator of the famous Hunterian Collection in London could instruct an emissary sent to investigate a stranded sea carcass, "It is of the utmost importance that the creature be found to belong to a species known to Science." The canon was closed. T. H. Huxley, who could tell bishops that he sat down "as a little child before a fact," yet wrote that he would not even investigate any psychic phenomena. The world of science was certain that it was open-minded and certainly it was not. There are few states of mind less favorable to real originality.

On the other hand, the churches were equally unhelpful. Determined not to own that their statements might need reinterpretation, they were discrediting the truths they held by refusing to distinguish between them and the forms in which earlier, less informed ages had expressed these truths. Thirdly, the practical businessman was not less certain that his way of thought and practice was not only superior to all his ancestors' but to all those of his contemporaries throughout the world if they failed to copy his habits and share his prejudices. A revolution was necessary if such a closed system was ever to open. It took place. We are not yet accustomed to think of the late nineteenth century as a revolutionary epoch, but it was. "Revolutions take place first in the human mind." Further, the whole modern age is revealing itself as a series of revolutions: First, the Religious Revolution, which began it; then the Political Revolution, which ended with the French Revolution; next the Economic, which broke out with Owen and St. Simon and culminated in '48 with Marx and the Communist Manifesto. These stages are now generally recognized. The fourth, however, being so near us, is only beginning to be detected—that is the Psychological Revolution.

It has been said that the real result of a reform is seen, not in the

pattern which the reformers substitute for the old way, but in the manner in which the old way proceeds after the reformers' departure; e. g., the counter-reformation church is in some respects more advanced and more different from the actual medieval church than are a number of the Protestant churches. Certainly after Darwinism and after the attacks of the Tübingen School and all the other higher critical studies had begun to be public, the further specific contribution of the churches was not latitudinarian liberalism or even the "social gospel." Rather, this contribution was shown in a return to an interest in the religion of direct experience, usually called mysticism. As Saudreau has shown *(Life of Union with God)*, throughout the nineteenth century and increasing rapidly in the last quarter, there was a growing interest in study and research in prayer, meditation, and contemplation, into the states of consciousness that attend such forms of attention, and the substantiated concomitants, psychical, psychophysical, and physical which such states produce. Here was a volume of evidence of contemporary psychical phenomena, and though the church as a whole was still anxious that this should not be exploited idly, there was a growing recognition of the need for critical studies of such states and their attendant factors. The simple triple division: divine, diabolic, delusory, all devout researchers realized, no longer covered the facts.

At the same time (as the former article on this subject has said) anthropology, itself a new science, reported evidence of para-normal experiences. Nor could these be dismissed as evidences of savage hysteria. The gradual translation of the Vedas and the Upanishads, the Pali texts and some Chinese Mahayana renderings and Taoist scriptures showed that such psychic phenomena were recognized and reproduced by men who in subtlety of thought and clarity of argument showed a mental ability of high philosophic caliber. In the biological sciences, after the spectacular success of natural selection as the explanation of life, a reaction began, raising anomalous facts which the popular theory had had to burke, while in pure physical research Maxwell's electromagnetic equations began to show to all mathematical thinkers that the human mind was dealing with a universe which was demonstrably extra-sensory. In psychology, although hypnotism remained suspect, Charcot was able to continue to arouse interest in the problems of the mind's power over the body, although

he considered that the capacity to be hypnotized was a mark, generally, of an hysteric temperament.

The mental climate of Western thought was, then, gradually becoming favorable to investigation mainly through the convergence of the two great schools of thought, the religious-philosophical and the mathematical-scientific. The former was prepared to consider at least having its contemporary evidence evaluated by critical method; the latter was prepared to consider that such evidence might exist.

Yet progress in this field might have been slower and more difficult than it has actually proved (and that would be saying a good deal) had there not been a third emergent factor to aid advance. William Bateson, the British geneticist who contributed largely to the revival of Mendel's work, made an interesting suggestion on this point. Being interested in history as well as in genetics, he proposed a correlation which is certainly suggestive. In common with all historical students he was impressed by the profound change in the mental climate which took place in Europe during the sixteenth century. Yet, in agreement with Lecky *(History of Morals)* and other authorities, he could not feel that one of the most excessive obsessions of the late medieval mind, its fear and hatred of witchcraft, was based wholly on an illusory phobia. Much of the evidence cited in these trials is extant. Nor was the procedure of many of the courts hasty, fanatical, or, if the corroborated evidence can be admitted at all, even unjust. The judgment arrived at was in accord with the evidence and with the laws of legal evidence. Though the punishments were barbaric, the legal process was often unexceptionable. Yet a few generations after such courts had been sitting as an everyday procedure, the whole of their *raison d'être* had become a matter of open doubt, contempt, and amazement.

What can be the explanation of such a revolution in opinion? Bateson's answer to this historical mystery was that of a geneticist. In his opinion, the witch trial courts had been anything but ineffective. In fact they had done their work so well that, owing to their capital methods, they had succeeded in doing what all other legal methods which use the death penalty have hoped to do but failed—they had exterminated the criminals. He felt that the only explanation which would cover the two facts—that there had been evidence and that, a few generations after, all such evidence had disappeared—must lie in

the fact that the courts had succeeded in exterminating all, or almost all, people who carried the capacity to manifest ESP. Naturally, when all such people were gone, the survivors, not experiencing among themselves such a queer gift, were convinced that earlier evidence for its existence must be fanciful. But, a few generations later, such stocks might well begin again to appear, especially in those distant areas, the Scotch and Irish highlands, where the Inquisition had never functioned effectively. Hence in the nineteenth century we should expect a steady increase in people carrying this faculty. So precisely when physical science was pushing beyond materialism, when religion was beginning to value increasingly the religion of immediate experience and to study the channels and methods through which that came, when Western man's assurance of himself as the final arbiter of knowledge was beginning to be shaken by his growing knowledge of other cultures and of their superiority to his in some psychological respects—at such a moment of new openness of mind, those who were ready to inquire were provided with subjects with whom they might begin fruitful study.

The more this series of convergences is studied, the more it would seem one of the most remarkable on record and the more it would seem to indicate that parapsychology is not merely a subject rapidly to become of general and informed interest but one which indicates the opening of a new epoch, an epoch when, largely through parapsychology and the successive studies it will promote, man may come to a new understanding of himself and of the universe which surrounds him.

THE ROLE MATHEMATICS HAS PLAYED IN ESP RESEARCH

By LT. J. A. GREENWOOD, U. S. N. R.

LORD KELVIN has said, "When you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." Though it has been difficult for many people to recognize, this is true of psychology just as it is of all the sciences. It is no less important for parapsychology, though some of those who are interested in the parapsychological branch have been slow to appreciate this fundamental necessity for the quantitative statement or evaluation of observations and experimental data.

In parapsychology, however, as in most other fields of study, people were interested in a range of spontaneous occurrences and unclassified observations long before any effort was made to apply mathematical aid to the investigation of them. Even after the founding of the Society for Psychical Research in 1882, which marks the beginning of more systematic investigation of the occurrences we call "parapsychological," there was on the whole relatively little use of quantitative treatment for many years. The main activity consisted in the collecting and sifting of reports of spontaneous experiences with only a small amount of mathematical evaluation applied to the classification and total findings. Even in the more experimental studies there was a considerable lag in the application of statistical treatment, probably because of the limited use in science in general of mathematical statistics at that period of scientific history.

From the present point of vantage, we may now look back over the history of parapsychology and say with reasonable assurance that the progress made has roughly paralleled the increasing application of mathematical method, that the more profitable period of scientific inquiry in the field began with the introduction of mathematics and statistics, and that the extension in the use of quantitative treatment

has been responsible for many of the greater gains that have been made and, consequently, for the increased recognition and confidence that have been given to the reports.

Owing to the fact that the very nature and conditions of the occurrences which have to be dealt with invite the counter-hypothesis of chance as an explanation, it is especially important to the study of parapsychological happenings that a clear-cut mathematical treatment be given. In fact, the classification of a phenomenon as parapsychical in nature requires a quantitative judgment as to its extra-chance character. The more explicit and precise this can be, the less is left to controversy. So definitely dependent is the evaluation of these phenomena of parapsychology upon all statistical techniques that we may almost say there would not have been a science of parapsychology without the development of probability mathematics and statistics. Instead, the chance hypothesis which apparently served for centuries to explain away the spontaneous occurrences of apparent parapsychical nature would have continued to be too powerful a competitor in the judgment of most critical minds.

Therefore it may be of interest, perhaps, not only to the student of mathematics but to the general student of parapsychology as well, to trace the main advances of mathematical techniques in the short historic period through which parapsychology has already gone. The parapsychologist of today can perhaps the better appreciate, from having such a survey, the difficulties which earlier investigators necessarily encountered, not only in conducting and interpreting their experimental work, but also in their efforts in presenting their findings to the critical attention of others. Something is always to be gained from the added perspective which the observation of general trends can bring us, more particularly in the greater alertness which comes from realizing that perhaps today we may still be handicapped, as were our predecessors, for lack of adequate instruments of analysis. We may the more quickly look for such instruments as a result of the historic perspective which shows the part they have played. And if nothing more is revealed than the close affiliation between the principles and methods of mathematics on the one hand and the problems of the real world as they appear in the sciences on the other—well, this alone is something which the student should never lose sight of.

* * * * * *

The first recorded investigation which was amenable to proper statistical or mathematical treatment was given by Richet in 1884 (22). He mixed several decks of ordinary playing cards, selected one card at random, and had the subject attempt to give the description of it. The reason this type of experiment is said to be more "scientific" than anything done formerly in this field is that his results could be reduced to a probability. That is, he could say how likely, or what the odds were, that a certain card would be named correctly by chance alone. With two colors, a correct choice would be expected, on the average, once in every two trials, and the probability is said to be one half. With four suits, the probability of chance hits is one fourth, and for a correct full description of a given card in a 52-card deck, the probability is one fifty-second. Every probability is a fraction, a number between 0 and 1. If the event has a high probability, a large fraction near 1, it is expected to occur relatively frequently. If it is very unlikely to occur, the fraction is very small and, if small enough, is accepted in statistics, for purposes of drawing scientific (or business) conclusions, as a practical certainty of the event in question not occurring.

The question immediately arose for Richet as to just what would constitute acceptable evidence of clairvoyance (or better-than-chance guessing) in the experiment in which he had ruled out sensory perception to his satisfaction. From past experience it would obviously be quite convincing if the subject guessed correctly all the cards presented over a considerable number of attempts. But it would still be convincing if only a few were missed. Just how many successes or hits should be registered, out of a given number of attempts, to be called acceptable evidence? A working answer to this query can be given if the results can be evaluated in terms of a probability—a probability that is small enough to be dismissed. Thus, in most fields of investigation, a probability of .01 or less (meaning that the results would be expected to occur by chance not oftener than once in a hundred times) is taken as a working level at which to conclude that something other than chance was at work to produce the results. Of course, the smaller the probability the stronger the conclusion, at least up to a point.

Richet's attempt to evaluate his data in terms of a probability was incorrect. However, the great utility of his work was that the data were

of such a nature that they invited mathematicians to produce correct methods of evaluation. In fact, shortly after his data were reported, Edgeworth published articles in the *Proceedings of the Society for Psychical Research* explaining correct evaluation procedures (4, 5). When the cards are returned after each guess and the deck reshuffled, it can be assumed that each guess has the same chance or probability of success as any other guess, entirely independently of it. Edgeworth pointed out in this case that the exact probability was a well-known algebraic binomial expansion which is generally too cumbersome to use. He showed that this could be approximated by the use of the "normal" curve, for which convenient tables were then extant, providing the results were expressed as follows: First, suppose there are n trials and the probability of a hit is p. (The probability of hitting the suit of a bridge card is ¼.) Let $q = 1-p$. Then $\sqrt{npq}$, called the standard deviation (SD) of the binomial case under discussion, is the unit of measurement to be used on the results. The expected chance number of successes is pn. The deviation of the observed number of hits from this chance value, expressed in units of SD, is called a "critical ratio" (CR) and the normal tables give the probability corresponding to a given CR.

This method of evaluating card-guessing results was used by a number of experimenters during the next fifty years and essentially without modification or supplementation. It has remained useful to this day, and it was not in fact until about 1937 that any question concerning its correctness or applicability to the uses made of it arose in the literature of parapsychology. However, with the publication of the period of the Duke work, quiet acceptance of the method of evaluation just described came to an end. Controversy flared up, and in the heat of it a number of charges were made that the mathematics used was incorrect or inapplicable.

It should be pointed out here that the story of the statistical issues with regard to the ESP work has been largely one of clarifying the points under discussion and, in some cases, of proving that the differences asserted were actually trivial for the purpose in view. But there are two points of all this critical discussion that do stand out as involving real mathematical issues and these will be reviewed.

In the first place, the ordinary deck of playing cards had been replaced by a deck composed of five squares, five stars, five crosses,

five wavy lines and five circles. Thus the probability of a single guess being correct was ⅕. But more important, Rhine had in general not replaced the card in the deck after each call or guess, but had allowed the subject or guesser to call all 25 cards without shuffling the deck between calls (though the subject was not told whether his calls were correct or not). Several people, notably Kellogg (15), questioned the appropriateness of treating this situation as though it were truly a binomial for evaluation of the results. (The procedure, however, was justified, as will be seen in the next few paragraphs.)

In this first genuine statistical issue to come into public prominence the question was, in effect, whether the subject could change the value of his scores by calling an equal number of each symbol to match the actual composition of the deck. This question could be pretty well answered by finding the actual expected value and standard deviation for this case. By elementary probability considerations, the expected value or mean was known to be *five hits per deck*, regardless of what sequence the subject called; that part was sound. Then the author published the variance (or standard deviation squared) and showed that the largest SD was to be obtained for this equal-calling or "matching" case (9). This called for a correction, as Kellogg had claimed, but the SD is *only two per cent* greater than for the binomial. This correction was recommended and was used until at a later point it was found to be actually unnecessary (though theoretically applicable as an outer bound) as will shortly be explained.

The probabilities of the different scores per run for the case just considered, where the subject supposedly called five of each of the five suits (called the matching case), had been approximated by Kellogg (16) and by Sterne (23). The exact values had been found by Huntington for the similar case with four suits (the 4 x 4 case) (14), and Greville computed them for the five-suit case (5 x 5 case), the one actually applying to the situation in the ESP test (or assumed to apply for purposes of discussion here). Greville extended his method to the case of an arbitrary set of calls for a deck in which an uneven distribution of cards per suit was given (12). The results showed the general applicability of the binomial method, as originally explained by Edgeworth, to the card-calling situation, as done by Rhine and others.

In some situations, an uneven deck has been used for the subjects

to call. Thus certain of the English experiments allowed the subject to ignore and pass over a card when he so desired. To allow for accidental coincidences of final effective deck composition and preference for certain symbols by the subject, a new mean and SD were needed. These, both Stevens (24) and Greville supplied. The author later published the third and fourth moments (average third and fourth powers of deviations of scores from the mean) which, with the mean and standard deviation, can give fairly accurate probabilities (10).

This took care of the theoretical side of the situation. However, the empirical investigation of it must not be discounted as it is the necessary "foot on the ground" connection between theory and fact. If then, ESP being ruled out for the moment, the experimental set-up under discussion produces the same results as if it had been to all intents and purposes a binomial situation, *and this sufficiently conclusively*, then it is fair enough to use binomial methods for evaluating the probabilities.

In order to settle this phase of the question, the author undertook in 1938 to match 100 runs of calls that had been made by subjects in earlier tests against 200 well-shuffled decks of ESP cards for which they were not intended. Each run consisted of 25 calls. This made a total of 20,000 decks called or 500,000 calls made in this chance check series. We know that had these all been made under the chance binomial situation, the run of 25 calls through a deck would have an average expectation of five hits, with chance distributing some higher and some lower. With 20,000 decks we further know the exact theoretical number that should exhibit 0, 1, 2, . . . 25 hits, respectively. The actual number of each obtained was counted, the theoretical and actual being shown in the table below.

By various tests, the observed and theoretical results are shown to differ insignificantly for the purposes at hand. Now it is well known that where any real fundamental difference of a theoretical and actual nature exists, then a sufficiently large series will show it up significantly. Since 20,000 runs is so many times larger than any of the experiments conducted to date and no real difference is found, it is logical to conclude that no real difference does exist within the range of a few thousand runs. And accordingly, the binomial affords the best available approximation to the test situation and the now tradi-

FREQUENCY DISTRIBUTION OF A 20,000 RUN ESP CHECK SERIES

Hits	*Observed*	*Binomial*
0	82	75.56
1	482	472.24
2	1,426	1,416.71
3	2,818	2,715.36
4	3,761	3,733.62
5	3,871	3,920.30
6	3,240	3,266.92
7	2,139	2,216.84
8	1,238	1,246.97
9	588	588.85
10	253	235.54
11	71	80.30
12	25	23.42
13	5	7.38
14	1	
15	0	
.	.	
.	.	
.	.	
Total	20,000	20,000.01

tional procedure of evaluation is approved *without need of correction* for the matching case.

* * * * * *

The second major problem was that of "optional stopping," i.e., the question of whether the results had been caused by stopping the tests at a point where, by chance alone, the deviation was significantly positive. The probabilities had so far been computed under the assumption that the experimenter had stopped the experiment on a principle that was independent of the results. Now if he were trying for, say, a large CR, the assertion was made that by computing the CR every now and then as the experiment progressed, the experimenter might by chance alone be able to come to a point at which he would have the large CR and there stop the experiment. Then, computing the probability on the usual basis he might announce significant results, and yet they would have come from what was really only a series of chance data. That this effect is actual is evident from the literature of probability, and a relatively simple proof of it for the binomial case was more recently given by Greenwood and Greville in 1939 (11). They proved that, theoretically, there is a probability of 1 (or certainty) that by continuing the series *sufficiently long* any preassigned CR could be attained. However, while this is an interesting point of theory, the important thing is to determine how effective

it really is in practice, and whether it has any bearing on the conclusions of the research in extra-sensory perception. Huxley's famous statement, "give a monkey a typewriter and plenty of time and it will reproduce Shakespeare," is equally correct theoretically but of no practical consequence whatsoever.

Again, there are two ways to attack this angle of the problem, the theoretical and the empirical. Theoretically, the optional stopping hypothesis could only be applied if the work as a whole gave evidence of being a chance series (from which parts could be selected, as it were, by stopping the experimenter at a "lucky" point). But, whether the results were pooled for all the various experiments for the first three years of the Duke work (thus wiping out all but the most recent stopping point) or whether they were presented as separate series, either way the results were found to be highly significant. There was no ground for supposing the data to have been from a chance series.

Again, from the theoretical viewpoint, both Greville and Greenwood devised safe ways of correcting for this effect which still permitted some freedom to the investigator. These are in the nature of giving larger probabilities to the results than the true ones, but not so much larger as to be impractical.

While some of the earlier experiments were apparently conducted without giving special thought to this question (which is rarely regarded in statistical work), during the past few years all series have either been corrected for optional stopping in the evaluations or else the experiments have been designed to exclude it as a possibility. Moreover, as stated, all of the earlier work at Duke was published in one general total pooling, which leaves no case for the optional stopping hypothesis. (For details of this argument, see *Extra-Sensory Perception after Sixty Years*.) For that matter, even after correcting for the optional stopping possibility, many of the earlier experiments still gave results that could not be reasonably explained by the chance hypothesis. The individual CR's were still significant, as was the CR of the total research on the side of practical empirical testing.

Considerable appreciation of the limitation of the optional stopping effect can be had from watching it in action in chance series. To this end, the large 20,000 run chance series previously mentioned was broken up into 400 blocks of 50 runs each (8). Each of the 400

blocks was investigated as to the optional stopping effect on the total running score of that block. *Only one attained a CR as large as* $+3$ (whereas large CR's have been conspicuous in the ESP reports). On the basis of the studies made, it was suggested in the article referred to that for a range of several hundred runs a CR of $+3.3$ might well correspond under optional stopping conditions to a CR of $+2.5$ without those conditions. Since a CR of less than 3.3 had been rare in the earlier ESP reports when the optional stopping hypothesis was not considered in the design of the experiment, it can be seen that the practical test just described affords a basis for acceptance of the work in question (i.e., on a $+3.3$ criterion).

* * * * * *

A third problem of a statistical nature—an old one, but one that has only recently been solved—is that of handling of drawings and non-quantitative material. This makes up a large portion of ESP data, but it had not been possible earlier to place a satisfactory probability valuation on it.

To some parapsychologists, the gain in exactness permitting probabilities to be assessed by the use of cards was accompanied by a highly probable loss of ESP efficiency. The tales of spontaneous parapsychic experiences were certainly a far cry from the calling of ESP cards. It was felt that the extra-sensory perception of meaningful things like portraits or drawings would strike nearer the fount of the supposed ESP capacity. Drawings had been used now and then for such experiments ever since any systematic approach had been made on the subject. But there was no satisfactory method of assessing the results once they were obtained.

The forerunner of the present methods for evaluating clairvoyant drawings, etc., goes back to a famous problem in probability called the "rencontre" problem. While it has been couched in diversified settings, they are all mathematically equivalent to the following version: n letters are written and their corresponding envelopes properly addressed. The envelopes are shuffled and assigned randomly, one to each letter. What is the probability that no letter will go in the proper envelope? Later generalizations asked for the probability that exactly r of the letters would go in their proper envelopes. The values essentially solving the problem were given by the famous Swiss mathematician, Euler, in the 18th century.

Psychologists had been using the principles of this matching of two sets of items for some years, with such items as handwriting specimens and personality sketches. Chapman gave the details of the mathematical evaluations necessary to reduce the results to a probability (2, 3). Pratt then made the first application of this technique to an ESP investigation (17). He evaluated some drawing experiments previously obtained by an ESP investigator, Dr. C. Hilton Rice.

Briefly, the data are obtained by presenting a set of original drawings or pictures to a subject or subjects who try to reproduce them with no clue as to what they are. Later, a judge attempts to pair them off according to similarities. The greater the number of correctly paired items, the smaller the probability on a chance basis. An essential feature of the use of the method has been that of utilizing the same data with several judges, thus achieving an economy of experimentation. The writer later pointed out that this procedure has no essential bearing on the alternative hypotheses to chance (6). So in this particular form the matching procedure has been discontinued in ESP research.

Carington has recently evaluated drawings by a method of "palpable hits" (1). The drawings are presented as previously described with the additional fact that the procedure is repeated with several sets of drawings. The measure of success which Carington uses is then dependent on the fact that more evident hits are made when a judge compares the subjects' drawings with their true originals than when he compares them with the sets for which they were not intended. The technical name of the test he uses is the "chi-square contingency test." He has obtained probabilities which are smaller than the value usually set for significance; that is, not acceptably explainable on a chance hypothesis.

Another type of drawings test and evaluation has been proposed by Stuart (25). It is called the method of "preferential matching" and is described briefly as follows: The originals are presented to the subject in order and he attempts to draw what they are or to put down something suggestive of them in the correct order. A judge (who may be the subject or experimenter if due precautions are taken not to disclose the correct order of the originals) then takes each of the drawings and ranks the originals with it in terms of likeness. The sum of the numbers or ranks assigned to the true correspondent in

the originals of each of the drawings is the score which measures the success of the experiment. Thus if 1 is assigned to the original most like the given drawing, then perfect success would give a score equal to the number of drawings made. Stuart has achieved notably significant results with this method.

The fact that the judges cannot be relied upon to assign rankings completely independently from drawing to drawing has produced a problem comparable to the issue between the matching hypothesis versus the binomial which was mentioned in an earlier section of this paper. The mathematical problem is in this case much more complicated, however, than that of independently assigned rankings and has required a special study for its solution.

In summary, it can be said that some considerable progress has been made, on the whole, in coping with the general problems of evaluating tests of free material, objects on which a definite probability cannot be placed as an *a priori* basis.

* * * * * *

It had been noted early that when hits in DT tests (see glossary) were plotted by position in the run of 25 trials, as total score for first trial, total score for second trial, etc., the resulting curves frequently showed the extremes (i.e., 1st, 2nd, 24th, 25th trial, etc.) deviating more from the expected value than did the trials in the middle portion of the run. In order to evaluate this effect in terms of a probability, Rhine suggested the use of a statistic which he called a "salience ratio" (or SR). This was a new application of a statistic which has been primarily useful in the analysis of variance; that is, the ratio of two chi-squares. In 1941, Rhine set up an experiment specifically to test the data for that sort of characteristic.

He further complicated the analysis of the results by introducing breaks in the run after every five trials, so as to give the effect of five small runs making up the one large run. He called these small runs of five trials "segments" and the 25 trials as a whole, the "run." He looked for relative success in the segment as well as in the run, compounding the segments of the run as if they were one, and getting five subtotals of hits within the segment. At the same time, he grouped the run distribution of hits into five subtotals, one for each segment. Then by taking the extremes divided by the three inner

positions, he obtained an SR which was either a segmental salience ratio (SSR) or a run salience ratio (RSR). Now, Rhine was looking for possible relations between the two sets of hit distributions as represented in the SSR and RSR, and required a means of comparing them with each other, as well as comparing the SSR's (or RSR's) themselves for the various subdivisions of the experimental data. The most generally used statistic to test the joint effect of two sets of paired items is called the "correlation coefficient." However, a basic assumption is that the two sets are each from a normal population. This assumption is clearly violated in the RSR and SSR data. So, to get around that difficulty, a comparable form of this called a "covariation statistic" was adapted by the writer to the needs of the situation described, one in which the higher moments play a part (7). Rhine's and Miss Humphrey's results in which this statistic plays a dominant part are published in a series of papers (13, 19, 20, 21), two of which involve the hypothesis of precognition.

The probabilities thus obtained for the covariation of the RSR's and SSR's assume that on a chance hypothesis the two measures applied to the same data are independent. The complete answer to the question: "Are they independent?" is not known. In view of that fact, Pratt, Humphrey, and Rhine conducted an extensive chance check and all the covariations computed in the usual way were found to be completely insignificant. We have, then, the fact that the chance results came out in good agreement with the assumed theory and the contrasting fact that the ESP tests with the same statistic came out in decided disagreement with it. It is a strong indication that whatever dependence there is, if any, is weak, and would therefore not have great effect on the computed probabilities.

* * * * * *

There are other problems that have been solved and other contributions of mathematics to the ESP research, but they do not seem to warrant the lengthening of this article to the degree their discussion would entail. Some readers will doubtless differ with the writer on the propriety of the selection represented here. There is bound to be some difference of opinion on these points. The introduction (by Jephson, with the aid of R. A. Fisher) of probabilities for the different cards of the playing card deck that would take preference into

account, as well as the empirical probability formula introduced by Saltmarsh and Soal for mediumistic material, are valuable leads which, in a fuller treatment, would merit sections in this survey were they less complicated to introduce. Both have left a trail of utility that has not even yet been abandoned. And there are other adaptations that have served for at least one important occasion. Fisher's chi-square combination method, the contingency table, and correlation techniques, are some of these further statistical tests which have found a place in the parapsychologist's range of instruments.

Perhaps it will not be taken amiss if this occasion is utilized for a few remarks on the general logic of application of the methods here under discussion; to many, they are, of course, elementary. Mathematics and statistics have taken many forms in the applications to ESP data, yet it should always be kept in mind that the final answer given by them to the query, "Did ESP account for any of the results of this particular experiment?" is only a probability that the data would have been obtained, assuming only the chance hypothesis to have been operative. Then, with proper exclusion of alternative hypotheses such as errors, etc., with a probability of the order of .01 or less, the experimenter may feel inclined to assert the existence of ESP in that experiment. That is, it can only give the probability that the observed data would be obtained *if* the chance hypothesis is the correct one. The experimenter then decides on the basis of mutual agreement, personal opinion, past experience, familiarity with the experimental set-up, analogy with other statistical investigations, and, to some extent, authority whether or not to reject the chance hypothesis and whether or not to accept a given alternative hypothesis as the explanation of the data he obtained. Statistics can never prove that the alternative hypothesis is the correct one to explain the data—it cannot even prove that the chance one is incorrect or insufficient. But on the other hand, it is in order to state how far the mathematical statistician *does* go in this matter, as well as where his function ceases, and the following statement may safely be made on that point: While specific instances of chance data may have been mistakenly assumed to have been caused in part by ESP (as must be inherently possible in any statistical situation), there is complete unanimity of agreement among statisticians that the entire set of series presented as evidence

of ESP did not arise under the chance assumptions by which they were evaluated.

Those who are familiar with the history of science need not be told that some of the greatest insights into natural science have had to wait for the invention of suitable mathematical techniques for handling the problems concerned, the most classic example, perhaps, being that of Newton's gravitation theory and the invention of the calculus. What we have seen in this survey of the mathematics of the ESP research and the way it has dealt with the problems that have arisen in that field, suggests that while there has been a much less dramatic association between the two aspects, the experimental and the mathematical, than perhaps has been true in certain more historic instances in other sciences, still the dependence of the experimental results upon mathematical analysis—and even newly adapted mathematical techniques—is hardly less obvious and essential.

What this survey shows most clearly, I think, is the need for a readiness in the experimenter, perhaps a greater readiness than has existed in the past, to turn to mathematical statistics to ask for special adaptations of methods or for the invention of new techniques of analysis to meet the problems that arise rather than to tolerate any lack of security in the interpretation of results or to neglect to wring from the experimental results the maximum of information they can be forced to disclose. It is to be hoped there will be no backing away from the investigation of a field of problems for lack of analytical procedures to handle the experimental results, not at least until mathematicians have thrown up their hands in despair over the possibility of providing the needed methods. Closer rapport between mathematician and the experimenter will presumably bring about the desired efficiency and the limiting factor will be shifted then to the limitations of the experimenter as experimenter and the mathematician as mathematician and will cease to lie between the two.

And finally, it is only fair to direct a comparable comment to those who play the important role of critics of research. There has been, particularly in the critical literature of the ESP research, a great amount of amateur and uninformed criticism of the mathematical aspects of the research. For the most part, this has been produced by persons not claiming to be mathematicians. It might be appropriate

to suggest here a more economical means of clarifying the mathematical issues. If before publishing a criticism of a mathematical point, one or more mathematicians with whose special field the problems are concerned are consulted, unsound criticism could be eliminated and much time saved. For just as the experimenter obviously needs the closest contact with the mathematical statistician in the course of planning and evaluating his experimental work, so the critic, too, has a like responsibility of checking up before making a hostile appearance in print and thus adding confusion to the already considerable task of presenting a new and difficult subject to a naturally hesitant professional audience.

References

1. Carington, W. Experiments with paranormal cognition of drawings. *J. Parapsychol.*, 1940, **4**, 1-117.
2. Chapman, Dwight. The generalized problem of correct matchings. *Annals of Math. Statistics*, 1935, **6**, 85-95.
3. ———. The statistics of the method of correct matchings. *Am. J. of Psychol.*, 1934, **46**, 287-298.
4. Edgeworth, F. Y. The calculus of probabilities applied to psychical research. *Proc. S. P. R.*, 1885, **3**, 190-199.
5. ———. The calculus of probabilities applied to psychical research. II. *Proc. S. P. R.*, 1886, **4**, 189-208.
6. Greenwood, J. A. A caution on the use of the method of correct matchings. *J. Psychol.*, 1940, **53**, 614-615.
7. ———. A covariation statistic. *J. Parapsychol.*, 1939, **3**, 163-166.
8. ———. An empirical investigation of some sampling problems. *J. Parapsychol.*, 1938, **2**, 222-230.
9. ———. Variance of the ESP call series. *J. Parapsychol.*, 1938, **2**, 60-64.
10. ———. Variance of a general matching problem. *Annals of Math. Statistics*, 1938, **9**, 56-59.
11. Greenwood, J. A., and Greville, T. N. E. On the probability of attaining a given standard deviation ratio in an infinite series of trials. *Annals of Math. Statistics*, 1939, **10**, 297-298.
12. Greville, T. N. E. The frequency distribution of a general matching problem. *Annals of Math. Statistics*, 1941, **12**, 350-354.

13. HUMPHREY, B. M., and RHINE, J. B. A confirmatory study of salience in precognition tests. *J. Parapsychol.*, 1942, **6**, 190-219.

14. HUNTINGTON, E. V. Is it chance or ESP? *Am. Scholar*, 1938, **7**, 201-210.

15. KELLOGG, C. E. Dr. J. B. Rhine and extra-sensory perception. *J. Abn. and Soc. Psychol.*, 1936, **31**, 190-193.

16. ———. New evidence (?) for extra-sensory perception. *Sci. Monthly*, 1937, **45**, 331-341.

17. PRATT, J. G. The work of Dr. C. Hilton Rice in extra-sensory perception. *J. Parapsychol.*, 1937, **1**, 239-259.

18. PRATT, J. G., HUMPHREY, BETTY M., and RHINE, J. B. A check on salience relations in ESP data. *J. Parapsychol.*, 1942, **6**, 44-51.

19. RHINE, J. B. Terminal salience in ESP performance. *J. Parapsychol.*, 1941, **5**, 183-244.

20. ———. Evidence of precognition in the covariation of salience ratios. *J. Parapsychol*, 1942, **6**, 111-143.

21. RHINE, J. B., and HUMPHREY, B. M. A transoceanic ESP experiment. *J. Parapsychol*, 1942, **6**, 52-74.

22. RICHET, CHARLES. La suggestion mentale et le calcul des probabilités. *Revue Philosophique*, Dec., 1884.

23. STERNE, T. E. The solution of a problem in probability. *Science*, 1937, **86**, 500-501.

24. STEVENS, W. L. Distribution of entries in a contingency table. *Annals of Eugenics*, 1938, **8**, 238-244.

25. STUART, C. E. An ESP test with drawings. *J. Parapsychol.*, 1942, **6**, 20-43.

A SINGLE SUBJECT IN A VARIETY OF ESP TEST CONDITIONS

By WILLIAM RUSSELL, U. S. Army, and J. B. RHINE*

ABSTRACT: After a period of informal testing for ESP ability involving a variety of test conditions and procedures, and with marked success in all of them, a young man was given four series of formal and carefully safeguarded tests in collaboration with the Parapsychology Laboratory of Duke University. One of these gave significant results, and the four series combined give a chi-square that is significant. The extra-chance character of the results is due to negative rather than positive deviations. Hypothetical explanations for this effect are discussed. In an incidental comparison of p-values (different probabilities of success per trial due to possible choices) ⅕ was highest in deviation, though not reliably so. In both the blind- and open-matching type tests V-shaped curves were found when success was plotted against key-card position. The contrast of score level between the Informal and Formal Tests is great and is the main topic of discussion of the paper. It is suggested to be due, not to increased safeguarding of experimental conditions, but to the change from a free, gamelike test situation in the Informal Tests to one of work and strain in the Formal.

INTRODUCTION

THE PRESENT status of ESP research is such that knowledge of some of the conditions affecting success in ESP tests is more valuable than further evidence of the phenomenon itself. So important is this question as to what favors or hinders ESP that often barely significant

* We acknowledge with gratitude the very cooperative spirit shown by the subject PH throughout the work here reported, and the assistance of his father Professor H. Their names are withheld to spare them possible publicity. The staff members of the Parapsychology Laboratory read the manuscript, and offered many suggestions for its improvement.

results, or those obtained under somewhat imperfect conditions, are worth consideration, especially in the formation of hypotheses for guidance in further exploration. This state of affairs is largely responsible for this report. Taken alone, the data presented here can scarcely be regarded as adequate for the establishment of the occurrence of ESP, but in conjunction with the available literature on the subject, this material contributes something toward an understanding of the process of ESP.

This study is based almost entirely upon the performance of a single subject, a youth of 15 years when the experiments began four years ago. He is the son of a university professor of mathematics who is a specialist in probability. The young man, now a sophomore in college, is intelligent and of good reputation. He often takes part in radio, newspaper, theater, bingo, and various other contests, frequently winning desirable prizes. His intense interest in ESP, psychic research, magic, novelties and games, coupled with much reading along these lines, was an important factor in the undertaking and promotion of these tests.

The subject, who will be hereafter referred to as PH, was initiated into the ESP tests by one of the authors, WR, who resided in the same city and was a friend of PH's family. From the beginning PH had success in working under a wide variety of informal conditions. After a series of successful performances, WR brought the case to the attention of the Parapsychology Laboratory at Duke and active cooperation with JBR was begun. The joint research which resulted continued through the summer of 1939, and after a lapse of two years was resumed when WR returned to PH's city in the summer of 1941.

It will be convenient to report the experiments conducted by WR alone under the heading of *Informal Tests* though by such identification we do not intend to convey the impression that they were incautiously conducted. They were not done with the use of the procedures and conditions adopted by the Parapsychology Laboratory. Under *Formal Tests* will be submitted all the work conducted as the joint enterprise of both WR and JBR with the exception of a short series of precognition tests, which, being connected with a distinctly different problem, was previously reported in the totals of precognition experiments by JBR.

The subject PH has also been investigated by Dr. C. E. Stuart (12) at the Parapsychology Laboratory. Significant results were obtained. There was also a short series of GESP tests carried out by Dr. Raleigh M. Drake in 1939 with PH. They were intended by Dr. Drake to be of exploratory nature only, with the hope that subsequent investigation would follow if the results seemed promising. The deviation was + 26 in 32 runs with a critical ratio of 2.30, which is suggestive although not significant.

Informal Tests

These tests are reviewed here to afford a basis of comparison in the discussion to follow of the effect of formalizing the test situation. If one refuses to consider any but the most carefully safeguarded experiments, he may pass over this section. Much has been learned, however, from experiments that are suggestive, even though not conclusive. The use of the Informal Test section will be in the direction of suggesting methods and conditions that may favor better performance in new tests with rigorous precautions which these in some instances lack.

Telepathy

Among the earliest tests tried were those in which PH acted as agent, and various friends as percipient. Agent and percipient were in the same room in all tests but one. PH stationed himself across the room with a composition book in which he recorded a number from 1 to 10, while the percipient, seated with an Ouija board, moved the indicator across the numbers thereon until he felt an impulse to stop. Usually 50 trials were made at a sitting, and between 17 and 23 hits were produced regardless of who was the percipient. Professor H, PH's father, agreed that the results were not chance, but raised the questions of unconscious whispering and similarity of thought. Therefore radio music was introduced to drown out any auditory leakage, and a deck of playing cards was used by the agent to choose the numbers to be sent (that is, the number was determined by the cut of the deck). The scores, however, were unaffected, and the experiments were repeated every few nights.

The pooled results of the tests with PH as agent which WR witnessed or took part in show 330 trials with 124 hits. This gives a

deviation of + 91 and a critical ratio of 16.6. Only chance results were obtained, however, when PH was eliminated and the percipients paired off with one another. Incidentally, all of the subjects except PH were skeptical of the occurrence of telepathy, although they all displayed interest.

On one occasion, PH chose alphabetical letters from a magazine while WR in the same room attempted to identify them, using the Ouija board. There were 21 hits in 52 trials, and in 13 of the 31 unsuccessful trials, a letter adjacent to the correct one was chosen.

In all the above tests, PH was agent. As percipient, he seldom scored above chance, but when tests of clairvoyance were tried his success was usually of the same magnitude as that of the percipients in the telepathy test when he was agent.

In the drawing tests conducted by Stuart (12) it is interesting to point out that the entire deviation of 47 points was secured in the five sessions where PH was agent, whereas the deviation with PH as percipient was zero. However, there were only two sittings under the latter condition.

Word Tests

PH next made efforts to determine with the aid of the Ouija board the first word on a designated page in a specified book. It was recognized that such attempts would not be easily evaluable, but the number of successes made it evident from the beginning that chance was not the determining factor.

Since in some instances PH chose both the book and the page number himself, the possibility of deception and unconscious memory is not completely excluded. Only the tests with better conditions will be discussed.

On March 5, 1939, PH left a book with Professor H and WR, asking that when he left the room WR should choose a page number, and that upon his return he would take the Ouija board and indicate on it both the page number and the first word on the page. WR placed the closed book on the piano, and chose page 33. Professor H suggested that his son could have heard this, and WR indicated page 46 by whispering. On returning PH took the board and obtained 4-7 as the page. He was then told the correct number, whereupon he

took the board and spelled out "f-a-i-t-h." When the book was opened the first word on page 46 was indeed "faith."

In April, when PH was asked for the first time to obtain the first word in the *second* column of a book, he was allowed to choose the page. While the book remained closed firmly in WR's hands he spelled "g-a-i-v-e-d," the correct answer being "gave."

The two books mentioned belonged to the H family; but on March 28 in his own home, WR went to the bookcase and picked a book of his own choice. The first two attempts by PH were failures, and then the word "i-t" was obtained correctly as the first word on page 51. A second book was chosen by WR, and again after two failures, PH spelled out the word "s-o-u-l," the correct word being "sound."

On the afternoon of March 5, PH obtained three consecutive words correctly, using the Ouija board in his own home, but concentrating on books in WR's library. PH reported the words over the telephone as "w-h-o," "w-a-t-c-h-d" and "c-l-a-s." When these books were examined the words were found to be "whom," "watched" and "class." Since PH was allowed to choose both the book and page numbers, the possibility of previous preparation by the subject is not excluded, but observation during handling of books in previous tests of the session cannot be the explanation. Immediately following this, PH was asked by telephone to change to another bookcase, whereupon he specified shelf two, book seven, page 43, "g-o-o-d." The word was found to be "grade" which has two letters right. Further efforts by PH were not successful.

Eventually, PH was asked to try word tests with books located in JBR's office at Duke, but of eight words tried, only one was correct, that being the relatively common word "t-h-e." Later, further attempts under these conditions were fruitless. At the same time PH became unable to read words from books placed in his own hands, and even when he was given complete access to them.

Card Tests

The first work done by PH with the ESP cards was unwitnessed and resembled the BT method. He recorded his calls in his composition book, and of a total of 1,160 trials done in January and February,

1939, there were 283 hits. This gives a deviation of + 51 and a critical ratio of 3.7.

In March, PH adopted the open matching (OM) method and was observed by WR. Five target cards were removed from the deck, and an effort was made to match the remaining twenty. This was done by distinct conscious choice rather than by an automatic manual response. Over a span of several days, 168 hits were produced in 440 trials, giving a deviation of + 80 and a critical ratio of 9.5.

As was pointed out at the time of their appearance, the commercial ESP cards are warped in printing in such a way that it is possible under certain lighting conditions to read the symbols from the backs. Therefore black opaque envelopes were obtained, and the cards enclosed in these. They were, however, left unsealed so that they could be checked immediately. WR shuffled the envelopes and handed them to PH, who matched them with WR present as usual. WR made the check-up or else PH was allowed to do so while being watched. Before each run the envelopes were reshuffled, and in several instances all cards were withdrawn and interchanged in the envelopes. Under these conditions, there were 58 hits in 130 trials. This gives a deviation of + 32 and a critical ratio of 7.0. The enclosure of the cards did not adversely affect the subject's performance.

After April, PH directed most of his ESP efforts toward various Formal Tests, using either enclosed cards or a compartment box which will be discussed later. Occasionally, however, the Informal Tests were again tried and often yielded brilliant results. On July 7, for example, although PH appeared to be scoring at a chance level while matching cards in unsealed black envelopes, when WR offered to take his picture with a camera, his scores climbed to new heights. In the entire session there were 43 hits in 110 trials, giving a deviation of + 21 and a critical ratio of 5.1.

During August, PH was observed by both WR and a newspaper reporter, HN. In a simple open matching test of 160 trials there were 57 hits, which gives a deviation of + 25 and a critical ratio of 4.9. HN, trying simultaneously, scored only 85 hits in 375 trials, giving a deviation of + 10 and a critical ratio of 1.3.

An incident that occurred at the time is worth mentioning. PH was asked to find via the Ouija board the color of the wall paper background just being put up in HN's home. The Ouija board spelled out "p-i..." and PH immediately said "pink," which according to HN was correct.

Single Card in Box

On the afternoon of July 23, 1939, in the midst of a series of formal tests by a different procedure, a perfect score of 25 hits in 25 trials was made with the ESP cards enclosed one at a time in a box. After PH had failed to better chance expectation in several runs of open matching and DT, WR produced a black bakelite box, measuring 6 x 3 x 2 inches with a hinged lid. There was a small hole as well as a crack in the box, but only if the box was inverted and held close to the eye could these be used as peep holes, and such peeping, as well as the inverting of the box, should have been detected by WR. Even so, this series would be omitted if it were not for the fact that PH succeeded only in one brilliant flash with this box, and thereafter failed to score so well again. This suggests to us that the slight openings were not used during the striking scoring, and therefore, as a suggestive case, this episode seems worth reporting.

A single card was chosen at random from the deck and placed in the box. With this procedure, the subject scored seven in a run of 25 trials. Then WR challenged PH to call them all correctly, reminding him how improbable it was from a mathematical viewpoint. He took a deck of colored ESP cards, shuffled it, and holding it to one side, pulled out a card, placed it rapidly face down in the box, and closed the lid. The subject then held the box, concentrated on the card, and named "square." When the box was opened it was found that the card was a square. Then the card was returned to the deck and the deck reshuffled. This process was repeated until 17 consecutive cards had been correctly named. Whereupon WR went into an adjoining room, closed the door, shuffled the cards, made the selection, and without looking at the card, placed it in the box and returned to the room in which the subject was waiting, leaving the deck of cards behind. Under these conditions, eight consecutive hits occurred, bringing the total to 25. The following is the sequence of calls and cards:

RSRSWPWPSWPPPWRCCRCWRWPWW

in which R, S, W, P, and C represent square, star, wave, plus, and circle respectively.

The fact that the last eight cards were put in the box in another room and the deck left there during the trial, leaves only ESP as an explanation of the subject's success, or the alternative possibility that the subject may have distracted WR's attention and peeped in the box. This is not likely since WR had the subject and box under close scrutiny throughout the experiment.

It is interesting to note that when WR attempted to continue the experiment after the 25th trial, the four further trials were all failures. In the first two of these, the card was placed in a book, and this being unsuccessful, a return was made to the box where the next two were likewise misses.

On September 7, an effort was made to repeat this experiment. A rubber band was placed around the box, and in some instances WR held his hands around it along with those of the subject. There were 11 hits in the 25 trials, for which the probability is about .006. In September, 1941, a further series of 88 trials produced only 21 hits against the mean chance expectation of 17.6, which is an insignificant result.

In brief summary, the Informal Tests have yielded unusual results, not ascribable to chance. If these results were due to ESP capacity, some evidence of this would be expected in the Formal Tests, provided these latter are such as not to inhibit its expression. The support of the Formal Tests to follow affords some plausibility, then, to the view that in the Informal Tests ESP ability may have been at work—and if it was, the success reported under the different conditions is of interest and value in the guidance of further research.

Formal Tests

On the basis of his performance on the Informal experiments, PH was given various tests designed in collaboration with the Duke Parapsychology Laboratory. This work is divided into four parts: (1) Enclosed Card Series, 1939; (2) Compartment Box Series, 1939, 1941; (3) Roll Box Series, 1939; and (4) *p*-Value Comparison Series, 1941. The procedures and results relative to these experiments will follow.

In these tests, PH did not repeat his earlier successes in getting

positive deviations, but instead showed a tendency to score below mean chance expectation. Either this change indicates inadequate safeguards on the earlier tests, or else a psychological effect produced in the subject by the more formal methods. The Formal Tests did, however, yield significant results, as will be seen below. After the results have been presented, the two alternatives will be discussed in relation to them.

Enclosed Card Series, 1939

Procedure. A number of decks of ESP cards, each card enclosed in a sealed opaque envelope, were prepared in the Parapsychology Laboratory at Duke and sent as needed to WR. In addition to ordinary sealing, each envelope was wrapped around with a broad gummed strip of paper which overlapped at the ends. The seal of the laboratory was then stamped over the gummed strip and the ends of the strip coated with shellac. These decks of enclosed cards were kept for the most part in the possession of WR and were examined, on their return to JBR, by two members of the laboratory staff. There were no grounds, either from the scores produced or from the condition of the envelopes, to suspect that they had been tampered with.

The subject matched the envelopes against five key cards representing the symbols of the ESP deck. After all the cards were placed opposite the key cards which they were supposed by the subject to match, they were turned over, and the code numbers on the backs recorded and sent to JBR for checking. At first no information was given to WR or to the subject as to the scores except the total deviations disregarding sign. Later when decks were returned along with the subject's calls, the information concerning scores and deviations was freely given as soon as the postal communication would permit. Duplicate records were kept by WR and JBR, and ultimately an independent check-up was made by WR.

Results. A total of 281.4 runs were made by PH in this series with a total number of hits of 1,447, which is only 40 above chance expectation. This gives a critical ratio of 1.17 which is not significant. There is, however, the interesting fact (pointed out here because of relation to items to appear below) that in the total of 35 runs in which PH worked alone and had of necessity to make his own record of

calls and code, he obtained a deviation of —25. This leaves the 246.4 runs conducted in the usual manner with a deviation of + 65. With this division of the results, the critical ratio for each condition was at once raised above 2, and the critical ratio of the difference is 2.7, which is significant. From the point of view, however, of the usual statistical evaluation, the results of the tests of PH given by this method are not presented as conclusive. Ten other subjects were investigated by WR on the same test during the same period. The comparison is slightly favorable to PH, but the difference is not of any statistical significance. The total of 425 runs for the 10 other subjects gave a total deviation from expectation of only + 10, compared to 281.4 runs with + 40 for PH.

Compartment Box Series, 1939, 1941

Procedure. The compartment box is an especially constructed card-board box measuring about 17 x 6 x 4 inches obtained by WR from the Parapsychology Laboratory. In the top of the box are five slits through which the cards may be inserted into the five compartments. The test is essentially blind matching since the task of the subject is to drop each given card face down into the compartment containing the key card which it matches. The key cards are placed in envelopes which are glued fast to the inside of the top of the box, one for each compartment.

It was agreed beforehand that ten "boxes" would be run (Series A), the subject inserting 750 cards, or 30 decks, into each. Five of these "boxes" utilized key cards enclosed in sealed opaque envelopes attached inside the box just under the top. After the subject inserted his cards, WR opened the boxes, recorded on serially numbered record sheets the number of cards of each symbol in each compartment, and sent the record to Duke for checking against the code of the sealed key cards. All scores were completely withheld from both WR and the subject. The five other boxes utilized enclosed but unsealed key cards. This permitted an immediate check-up which was made independently by WR and Professor H. Thus the experiment was designed to study the effect of delay or withholding information regarding the scores.

Before each "box" was run, the decks were manually shuffled for a half-hour, in order to randomize the cards. A variety of methods

was used. The subject then required one hour to distribute the cards into the compartments of the box.

In the immediate check series, five ESP cards were placed inside the black opaque envelopes inside the box just under the top and clipped, but not sealed. This was done by Professor H in a closed room. The box was then sealed with wax and presented to the subject who proceeded to drop in the cards in WR's presence. Afterward, Professor H and WR independently counted the cards, looked at the keys, filled out the record sheets and counted the scores. Great care was exercised, and no discrepancies between the independent check records were found. The subject was then called in and told his score. All record sheets, both used and unused, were returned to the Parapsychology Laboratory at Duke.

Results. There were 7,350 cards, or the equivalent of 294 decks, dropped into the compartment boxes in the first round. (One "box" consisted of 24 instead of the usual 30 decks.) The results, as shown in Table 1, gave a negative deviation of 92, which has a critical ratio of 2.7. The critical ratio of the difference between the delayed check and the immediate is negligible (CR = 0.4).

Table 1

COMPARTMENT BOX, SERIES A

Dates	Conditions	Runs of 25	Dev.	CR
7/13/39 - 8/9/39...	Delayed check and scores withheld. Sealed key cards.....	144	−52	...
7/18/39 - 8/15/39..	Immediate check. Enclosed, unsealed key cards.....	150	−40	...
Total Series A.....	.....	294	−92	2.7

Next it was decided that ten additional "boxes" would be run (Series B), five in the presence of WR and five by the subject alone. Two new sets of five sealed key cards were obtained from Duke, and all work was of the delayed check type with scores withheld.

In this type of test, the question arises as to whether the subject might not have obtained sensory cues from the packs of cards or, on the five boxes on which the subject worked alone, might deliberately have looked at the cards. This might lead to placing a large number of the cards of one symbol in a given compartment, giving a "stacking

effect." Such an effect would be detected in the checking and the record would show an excessive number of symbols of one kind in each compartment. Two boxes were found with records that could have been produced by stacking (though we cannot say that they were), and these were at once ruled out as potentially misleading in some of the analyses to be made. They comprise 60 runs with a deviation of —7 and their elimination leaves the critical ratio of the 1939 series exactly unchanged at 2.79, and the final critical ratio of all the compartment box tests unchanged at 2.83. Hence there is no need for further discussion of the two deleted box records.

With these deletions, the total of 240 decks in Series B gives a negative deviation of 37 and a critical ratio of 1.2 (Table 2). When pooled with Table 1 results, the total for the A and B series is 534 decks with a deviation of —129 and the critical ratio is 2.79.

Table 2

COMPARTMENT BOX, SERIES B

Dates	Conditions	Runs of 25	Dev.	CR
8/18/39 - 8/31/39..	In the presence of WR. Sealed key cards. Scores withheld....	150	−19	...
8/22/39 - 9/2/39...	By subject alone. Sealed key cards. Scores withheld.....	90	−18	...
Total Series B		240	−37	1.2
Total Series A and B		534	−129	2.79

But the question arises as to whether there might not still be stacking in the other boxes that was not obvious by inspection. This would not be expected to contribute to the deviation, but might require a larger standard deviation than $\sqrt{npq}$ because of the increased variance.

Accordingly, an unbiased estimate of the variance was calculated, and the standard deviation obtained by taking the square root. Fifteen cross checks gave a SD of 58.5 while the actual box scores yielded 39.8. The theoretical SD obtained from $\sqrt{npq}$ is 46.2. Certainly the SD obtained from the actual scores is on the safe side, and if there is a hypothetical stacking effect in the cross checks, it may have been produced quite as well by misdirected ESP as by sensory cues from the cards. Especially is this so when we consider that the compartment box method is essentially a blind matching technique, and a

subject could easily form a conscious or unconscious idea of what one or more of the key target cards were and misdirect some of his ESP toward matching the other cards accordingly. Moreover, since the deviations are negative instead of positive, it is possible too that the "stacking" may have been due merely to the ESP effort (perhaps unconscious) to place the cards in the wrong compartments. These alternatives should at least be recognized as possibilities, along with the stacking hypothesis.

In any case, another series of compartment box tests was run during the Christmas vacation period of 1941. Seven boxes were tried (there was not sufficient time for more), and the experiment was so planned that not only were there five sets of key target cards, enclosed, sealed, and attached within the box, but all 500 cards (20 decks) used at each sitting were each enclosed in a black opaque envelope and well shuffled. These envelopes were numbered on the side away from the subject as he inserted them into the slits of the box. This necessitated the labor of recording the number of every card and the compartment into which it was placed, and ultimately making a code and totaling the number of cards of each symbol placed in each compartment. The deviation of the 140 runs was — 18 which is not significant, although in the same direction as the other boxes.

In spite of this extra precaution to prevent any stacking, the SD obtained from an unbiased estimate of the variance is 27.5 for the actual box scores, and 25.7 for ten cross checks. The theoretical SD is 23.7. Since the observed values are again slightly above the theoretical, and yet the experiment well planned against any sensory leakage, it seems quite legitimate to use the theoretical standard deviation rather than one obtained from cross checks and influenced by symbol distribution inequality, which *may* be due to a slight stacking effect, but is more likely due to the misdirected ESP instead.

The additional 140 runs may be considered as a C series and if added in, give a total of 674 runs for all the compartment box work done by PH, with a total deviation of —147 and a critical ratio of 2.83. In any case, the results taken in their entirety are significant, even when the correction is made for deviation in both directions. For the A and B series $2p = .0054$, and for A, B, and C $2p = .0048$. The optional stopping question is not of too much concern since the

first series (A) alone gave a critical ratio of about the same magnitude (2.7).

No significant differences are shown either between delayed and immediate check-up, or between the subject working alone and in the presence of WR.

Three other subjects who had minor success in qualifying tests also tried compartment box experiments with sealed key cards. Their results are of interest for comparison. The total of 290 decks gave a deviation of —24 which is insignificant and represents a smaller rate of deviation than that of PH.

It is interesting to observe the consistency of the scoring below the expected mean; that is, the preponderance of negative deviations throughout. Not only are the three main series negative, but the two subdivisions of Series A, also, as well as those of B. This amounts to five negative groups. In addition, the results discarded for possible stacking, which would not affect the sign, had a negative deviation and the additional work by three minor subjects gave a minus deviation too. All the seven incidental subdivisions, then, are negative. That they would be all of the same sign (either + or —) is to be expected about once in 64 such total sets of data. This is independent of the significance of the *size* of the deviation.

Roll Box Series, 1939

Procedure. The roll box is a metal box measuring about 10 x 8 x 4 inches. It has on its top surface an opening through which the subject can write a row of 25 symbols on a roll of paper enclosed in the box. The objective is to match the row of target symbols which are printed on the roll and are at a designated position, but invisible, under the metal top. Once the subject has recorded his trials, the roll can be turned by means of a knob so that the record moves under a glass window where it can be seen but not changed. The target row symbols at the same time move into the opening. In this way, the two rows can be checked, one against the other, and the number of hits determined immediately.

This box was loaded and securely locked at the Parapsychology Laboratory, and sent out to WR. PH then made 118 runs yielding a deviation of + 9 and a critical ratio of 0.4 which is not significant.

p-Value Comparison Series, 1941

Procedure. In August and September, 1941, PH was tried on some simple informal tests, and he again averaged above expectation, though not as well as in 1939. Accordingly, further enclosed tests were tried. This time the problem for investigation was not merely to obtain evidence for ESP performance, but to determine which p-value was most favorable for success in testing. (By p-value is meant the probability of success on a given trial, which depends upon the number of possible choices.)

The experiment was planned by WR and at his request the following decks of cards were made up in the Parapsychology Laboratory and mailed to him, each card enclosed in a sealed opaque envelope: one deck of playing cards, one deck of ESP cards, and two decks of 25 cards, each card bearing the typewritten name of a popular movie star, prominent international figure, or famous man of history. In addition to these, there was one deck of 24 cards made up by Professor H consisting of 8 crosses, 8 circles, and 8 blank cards. By using the playing cards for calls of color, suit, rank, and complete identification, p-values of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{13}$ and $\frac{1}{52}$* were obtained. The deck of circles, crosses, and blanks gave a p-value of $\frac{1}{3}$, the names deck, $\frac{1}{25}$, and the ESP deck, $\frac{1}{5}$. Because of a variety of circumstances, it was not convenient to get an equal number of trials for each p-value. Likewise, the difference in materials might have produced inequalities in performance in the different decks. In any case, the experiment was an exploratory one. The records were made by WR, using code numbers placed on the back of the cards for identification; they were checked by JBR and later rechecked by WR.

Results. Table 3 gives the number of trials, the deviation and the Foster ESP quotient (1) for each type of material and p-value. While no independent significance was attained, it is interesting to note that the distribution is similar to that obtained by summarizing earlier reports presented in *Extra-Sensory Perception After Sixty Years*, page 271, in which the p-value of $\frac{1}{5}$ gave the highest ESP quotient.

* This, of course, involves a double call, as "10 of hearts," simply combining a p-value of $\frac{1}{13}$ and $\frac{1}{4}$ respectively.

Table 3

P-VALUE COMPARISON—PH—1941

Material	p-value	Trials	Dev.	ESP Q*
Playing cards (color)	1/2	312	− 7	− 4.5
Cross, circle, blank	1/3	216	+ 6	+ 4.2
Playing cards (suit)	1/4	416	− 6	− 5.8
ESP Cards	1/5	1,600	−35	−10.9
Playing cards (rank)	1/13	780	− 5	− 8.3
Names	1/25	1,300	− 2	− 3.8
Playing cards (suit and rank)	1/52	416	+ 1	+ 0.3
Totals	....	5,040	−48 SD†= 23.94 CR = 2.00	...

* See Foster, this JOURNAL, 1940, **4**, 320.

$$\text{ESP Q} = \frac{100d}{nq} \text{ for + and } \frac{100d}{np} \text{ for − deviation.}$$

† S.D. $= \sqrt{\Sigma\,(\text{S.D.})^2} = 23.94$ CR $= 2.00$.

The Formal Test Results Combined

One of the four Formal Test Series (the Compartment Box) gave an independently significant result; another, the p-Value Comparison, gave a result that is suggestive. When the critical ratios of these are combined by the chi-square method and the other critical ratios, of course, included, a chi-square of 13.56 is obtained (see Table 4), and this with 4 degrees of freedom has a probability of .009 which is significant.

Table 4

RESULTS OF THE FORMAL TESTS SUMMARIZED

Test	Trials	CR	X^2
1. Enclosed card	7,035	1.17	1.37
2. Compartment box	16,850	2.83	8.01
3. Roll box	2,950	0.42	0.18
4. p-Value comparison	5,040	2.00	4.00
			13.56
			4 d.f. P= .009

Relation between Success and Key Card Position

Another reason for believing that something extra-chance was occurring in the Formal Tests is found in an effect noted by WR in looking for a relationship between success and key card position. It is interesting to note that when the Compartment Box records, the

Enclosed Card records, and the p-Value Comparison records are examined, V-shaped curves are obtained in almost every case, when the key card position is plotted against the success ratio. That is to say that PH scored more negatively on the inner target card than on the outer target cards, as is revealed by the following Table:

Table 5

RELATION BETWEEN SUCCESS AND KEY CARD POSITION—SUBJECT PH

Test	I	II	III	IV	V	Total
Compartment Box A, B*, and C						
Trials	2,129	1,851	4,801	3,612	4,457	16,850
Deviations	−11.8	+ 9.8	−93.2	−31.4	−20.4	−147
Per cent Success	19.4	20.5	18.1	19.1	19.5	19.2
Enclosed Cards OM, 1939						
Trials	1,174	1,454	1,410	1,562	1,435	7,035
Deviations	+ 9.2	+14.2	− 5.0	+12.6	+ 9.0	+40
Per cent Success	20.8	21.0	19.6	20.8	20.6	20.6
p-Value Comparison OM, 1941 (p=⅕)†						
Trials	102	130	143	146	129	650
Deviations	− 1.4	0	− 8.6	− 9.2	− 5.8	−25
Per cent Success	18.6	20.0	14.0	13.7	15.5	16.1

*Incidentally, the two deleted stacked boxes also give a V-shaped curve with positive deviations at positions 1, 2, and 5 and negative deviations at 3 and 4. This deletion therefore does not affect the results nor the conclusions relative to the V-curve effect.

†Only 26 of the 64 runs (1,600 trials) of p-value 1/5 were done by the OM method. The remaining 38 runs were done by a calling method as were also most of the runs of the other p-value tests. However, 416 trials with p=1/13 were done by OM and also give a slight V-shaped form. Matching with p=1/2 obviously cannot be used in the study.

The accompanying graph (Fig. 1) shows the relationship more clearly, and it appears therein that the three series give a relatively V-shaped curve. Since such curves have not been previously reported in matching techniques, it may be best to consider their occurrence in these data as suggestive only. However, they are of interest both as an actual effect and as an indication that the Formal Tests do contain ESP.

DISCUSSION

Is There Evidence of ESP?

Before it is profitable to enter into a discussion of the conditions that might have affected the production of the results, it is necessary to discover whether or not it is reasonably probable that PH has shown ESP capacity in some portion of the tests taken. This may

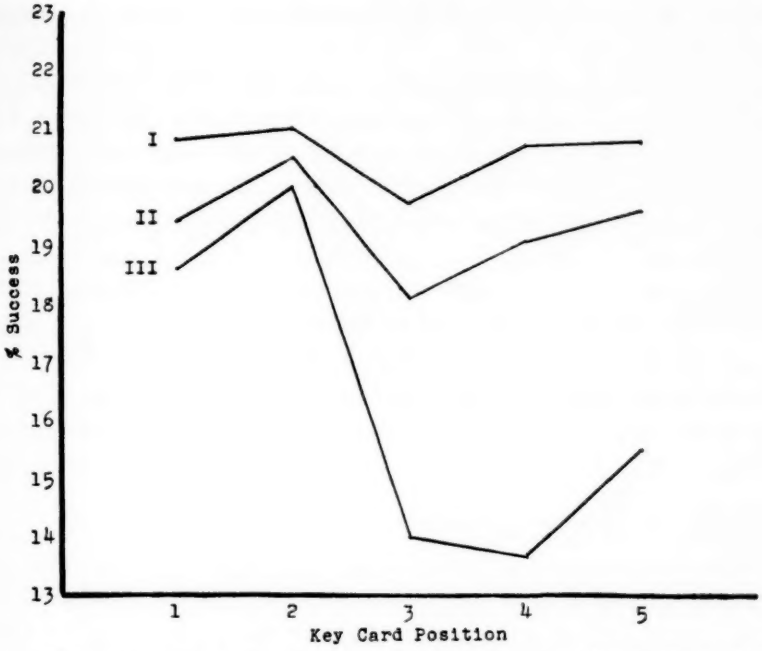

FIG. 1—PH's scoring success in relation to position of key cards. I. OM, 1939; II. Compartment Box, BM, 1939, 1941; III. p = ⅕, OM, 1941.

best be accomplished by a brief review of the more common alternative hypotheses and a discussion of the conditions provided for their exclusion.

The Informal Tests were extra-chance, but there will be no attempt made to utilize them as a basis for decision in the primary question of the occurrence of ESP.

The largest of the four series of Formal Tests gave a significant critical ratio of 2.83 while the chi-square combination of critical ratios of the entire work gave a marginally significant probability of .009. There is, therefore, a moderately good assurance that the work of PH in the Formal Tests as a whole is not ascribable to chance. This position derives no little support from the independently reported Stuart study of PH's ESP capacity. Had this work been incorporated in the present report, it would have belonged in the Formal Tests group and the combination chi-square would have been *more* than marginally significant. In view of this and the evidence as a whole.

the case is regarded as sufficiently established against the chance hypothesis.

The conditions with the Formal Tests were such that the subject could not possibly have obtained any sensory cues whatever. The Compartment Box Series, which is independently significant, utilized nine sets of key cards which were never in the subject's hands and which were securely sealed and later opened at the Parapsychology Laboratory.

The plan of independent checking and computation reduces the likelihood of error to a point of unimportance.

With the results of the Formal Tests acceptably significant, it is justifiable, particularly in the search for knowledge of the effect of test conditions upon ESP, to give some attention to the Informal Tests.

Negative and Positive Deviations

The first negative deviation of importance appeared after the institution of the Formal Tests. When the results of the Enclosed Card Series were divided according to whether the subject worked alone or in the presence of WR, it was found that the part of the series with the subject working alone gave a negative deviation (CR = 2.1). Later, the compartment box was found to yield significantly negative results. Finally, there was again an almost significant negative deviation in the Sealed Card Series of 1941. The Formal Tests were, therefore, preponderately of the negative deviation type, while the Informal Tests were all positive. The contrast is of more than ordinary interest.

These negative deviations are quite as interesting as equivalent positive ones would be. For a subject to score significantly below chance implies either a knowledge of the cards or something equal in effect. Much discussion of the cause and *modus operandi* of negative deviations has already appeared in the literature (2, 4, 5, 7). In general the attempts at explanation involve either a supposition of unconscious reversal of motivation (negativism) on the part of the subject, according to which the symbols are called something else than what they really are in order to thwart the experimenter, or else a more cognitive explanation is proposed by which the subject unconsciously distorts his aim and therefore blocks the correct symbol by

his very efforts to apprehend the right one. (The simile of blindly maladjusted sights on a rifle affords a useful parallel.) It will be sufficient, however, in this instance simply to describe some of the conditions which we think are relevant and leave to a later period an attempt at the actual explanation—a period when future research has added much more to our knowledge of the workings of ESP.

The general effect of the transfer of PH to the Formal Tests, and, as it turned out, to negative deviations, was to go from a game-playing status to one of work. Up to the period when connections with the Parapsychology Laboratory were established, the tests were all in the spirit of fun and of display to friends. There was at most an air of amateur inquiry. The introduction of the Duke contact naturally brought a sense of responsibility and the recognition of the necessity of making good on the basis of claims that had been made. In addition it was now necessary to make out records of calls and code numbers, and the check-ups of these involved considerable delay. When PH did not participate in this labor, it seems not to have depressed his score below mean chance expectation, though it did depress it far below his Informal Test level. When, however, he was left alone by WR and had to do the call and code recording, there seems to have been something in the situation that induced a negative deviation.

In the instance of the 1941 series, where he went negative even with WR present, we can only suggest that PH had long since become satiated with this type of procedure. It was new only in the variety of material and p-values, and since most of the tests were done with ESP cards, there was not even at the beginning any variety in this. Our suggestion would be, then, that some internal resistance due to boredom might have resulted from the nature of the test. This is to some extent borne out by the fact that in the series of tests conducted a little later by Dr. C. E. Stuart of the Parapsychology Laboratory (12), tests which involved considerable novelty for him, PH returned to significant positive scoring.

Finally, in the instance of the compartment box, the circumstances we would most likely associate with the negative deviation are involved in the fact that the subject had before him the long task of depositing 750 cards through the openings of the box. Then there was either the delay of checking the records sent back to Duke, or the

delay, inconvenience, and formality of the routine check-up by WR and Professor H.

In exchange for added safeguarding in all these Formal Tests, there was lost the enlivening prospect of immediately available information of success at the close of each run, as had been the rule in the Informal Tests. Equally absent was the opportunity for PH to turn, when the least bit bored, to some other variety of test, or to explore whatever innovation came to mind. There is, of course, no finality to the connections suggested between conditions and scoring trends. They are mentioned here to add to the body of knowledge of conditions associated with negative deviation with the hope that they add something of value to the urgent need to discover the determinants of ESP scores.

p-Values

There is little to add to the facts themselves, as given in Table 3, regarding the p-Value Comparison Series. This, like the summary referred to in *Extra-Sensory Perception After Sixty Years*, would appear to show ⅕ to be the most successful p-value when the results are measured by the ESP quotient. There is still a possibility that the results might be an effect of PH's increased familiarity or satiation with the ESP cards or of the type of symbols used rather than the number of choices possible. The object in choosing the ESP symbols had been to avoid confusion between symbols, at the same time attempting to get the simplest design possible. The number five was a compromise between the need for variety and the need to limit the number of symbols to one easily retained in the spotlight of consciousness. The suggestive findings of this report and the earlier work can be claimed only to accentuate the need for thorough investigation, one in which the controls will be adequate to narrow the conclusion to the point of issue. This will probably mean using a common stock of symbols for all the p-values, equalizing also the number of calls for each symbol and for each p-value.

Another point of importance is the use of the ESP quotient as described by Foster (1) for the measurement of ESP in negative deviations. The propriety of the usage depends upon whether equal knowledge is shown by negative and positive deviations the same size, or whether, as is implied by Foster's formulation, four times as much

knowledge is necessary for a given negative deviation as for an equal positive one when $p = \frac{1}{5}$. This matter has been discussed at length by Rhine (5) in an earlier paper, but the issue is not settled as yet. But even on the basis of percentage of trials, the deviation for a p-value of $\frac{1}{5}$ is the highest and this ignores the direction of deviation.

The Use of Cards

There is one point of interest that can be profitably taken from the Informal Tests above, assuming that they probably represent the operation of ESP. There has always been discussion as to whether card tests are as good as a variety of other tests that have been compared to them. Drawings have been much used and argued about as an alternative to the card method, and indeed much excellent work has been done with them. However, it is difficult to find anything in the literature of ESP more convincing so far as the counter-hypothesis of chance is concerned than a run of 25 consecutive hits. The half-dozen instances when this has occurred are scattered, and information regarding them is not assembled in any one place; nor in some cases are the conditions exactly what they should be. But since we are speaking of the type of material, it need only be said that few experiments have produced anything similar which would have such an impact on one's judgment as a perfect run through a deck of cards under adequately safeguarded conditions.

By this we do not mean that ESP material should be limited to card calling. On the contrary we favor a wider variety of test materials than has been used. The point is merely that PH was as successful on the cards as on anything else tried.

The Question of Decline in Scoring with Improvement of Conditions

The first impression given by these results on the transfer of PH to the Formal Tests is that the mere improvement of conditions themselves brings about a decline in scoring level. The trouble with this impression is that it fails to distinguish between the various effects which the transfer to the Formal Tests might have upon the subject. It is obvious that a great deal more is happening than the introduction of safeguarding; for example, the enclosing of the card in a sealed opaque envelope does more than exclude vision; it makes it impossible to check up on the subject's run without long delay, it places the sub-

ject under some increased strain, it changes the use of such cards from an informal gamelike procedure to the handling of experimental materials with the laboratory atmosphere. When we consider the importance of freedom and spontaneity in similarly delicate mental processes such as humor, worship, contemplation, reverie, and the like, we can readily appreciate what the intruding hand of precaution may be doing to the sensitive process under investigation.

This question has been discussed at greater length in *Extra-Sensory Perception After Sixty Years*, pp. 198 ff and 236-7. But in the present discussion, the question is referable to the actual statistical results and to the adequacy of the experimental conditions. If, in spite of all precautions to eliminate error, the results still indicate that something beyond chance occurred, the reader is left the alternative either to suppose that there is still another source of error not discovered, or that ESP is operating. We, along with the other experimenters in the literature referred to, are unable to find another source of error in these experiments, and if ESP occurs under the more favorable conditions, we think it much more reasonable to suppose that the introduction of the Formal Tests acted as an inhibition on that process. The drop in scoring, not only to chance but to a point below it, to the extent of attaining statistical significance through negative deviation, strongly suggests that psychological effects of the Formal Tests have been responsible for the decline, rather than the prevention of hypothetical error by imposing more controlled conditions.

It is, of course, the hope of every experimenter that better test procedures and conditions will be discovered, and by means of these that better accommodation jointly for the freedom of the subject and for the security of the conditions can be found.

Conditions Favoring Success

The discussion of different p-values above would belong under the general question of conditions favoring success as would also the discussion of the use of cards as opposed to other materials. To these might be added distinctions between the roll box and the compartment box, the one showing only chance results and the other a significantly negative deviation. But all these topics of discussion concern the type of test or test material as such, and important as these

are, they are apparently not the most important conditions determining the success of the ESP test.

If we consider again the results of the Informal Tests, and suppose for the moment that they have to do with ESP, we see at once that the conditions favorable to success are to be found in that group of tests rather than in the Formal. Obviously, if the discussion of these conditions and their relations to performance could furnish any insight that would be applicable to future testing in a practical way, the resulting progress would be remarkable. And while we cannot do more than hope for such an outcome in the present instance, such a discussion may be profitable.

We have already referred to what may be the most important difference between the Formal and Informal Tests; namely, the transition from a casual, gamelike situation to one in which the subject has a task assigned to him, usually with the prospect of a fixed amount of work to be done, lasting for many days. There was also the subject's need to make good on the claims that had been made for him by WR. Readers who are aware of the inroads of stage fright upon the delicate mental processes will have no difficulty in imagining some of the inhibiting stress that PH probably experienced in calling cards in sealed envelopes on demand, or using compartment boxes sent out from Duke University.

It seems very probable that the mere connection with a university experiment may, in itself, give the lay subject some especial feeling of exaggerated strain, accompanied by a feeling of fear of failing in what is expected of him. It is, of course, speculative to assume that such a state of mind is responsible for the lowered scoring of the Formal Test series, but this is something that can presumably be found out in a general way; experimentally, at least, it should be possible to determine whether subjects do better when they are free from any responsibility of having to make good a claim made for them, when they are relieved of any fear of making a bad impression, when there is no sense of obligation or exaggerated responsibility about the task, and yet are eager to succeed.

Another point of contrast between the Formal and Informal Tests is the spontaneity of interest on the part of the subject himself. In the Formal Tests, the subject's tasks were laid down for him, while in the Informal Tests for the most part the type of work done was of

the subject's choosing, or at least he had the power to veto it. Furthermore, the subject had the opportunity to stop at any point he wished and to change to another method. He could select his partner for a telepathy test and have a great deal to say about the conditions or at least his acceptance of them. All this was reversed with the introduction of the Formal Tests; then it was WR, and JBR by proxy, who were dictating conditions and rules. It is true, an effort was made by the introduction of rewards and, on the part of WR, by the variation, insofar as possible, of the unessential conditions, to foster an atmosphere of freedom and spontaneity. But in the face of the handicaps this did not succeed so far as could be judged.

In the Informal Tests, there was an important factor inherent in the personal relation. PH and WR were trying the experiments simply because they were both interested. Presumably PH wanted to make an impression on WR (and on other persons in touch with the tests; e.g., Professor H) and much of the zest for the performance came from having an audience to impress. In the Formal Tests, on the other hand, while WR was present most of the time, he could not be a responsive audience because he had no knowledge of PH's success at the time. There was no spice of excitement over a high score, even when it occurred, because it was not known to anyone present. There was, in effect, no appreciative reaction from persons present. Although the challenge was still there to achieve high scores it was a relatively abstract one.

It is possible, of course, that even though WR did not know the scores in the Formal Tests series, his presence gave some stimulation to PH. This should at least be expected since the two had been associated throughout in PH's ESP work. This is suggested by the Enclosed Card series, when, during 35 runs of the series which gave a negative deviation of 25, WR left the subject alone and PH had to do his own recording. Up to that point, he had been scoring above expectation, and it is at least suggested that the negative deviation was due to the changed social situation. Since there is a significant difference between the runs obtained in the series with and without the presence of WR, some conviction is warranted on this point.

The impression was left distinctly on WR during the Informal Tests that a certain amount of light and good-humored challenging had a beneficent effect upon PH in the tests. This reached its climax

in the single-card-in-the-box test, in which a run of perfect scoring was obtained. WR played up the aspect of challenge and the excitement of the subject reached a high pitch during this experiment. Since that time, the continual and liberal use of rewards and challenges has had little or no effect.

It is easy to understand that the motivation built up during such a run has a natural breaking point in the climax of the 25 trials, and that the failures would have followed as they did. It will be recalled that Rhine's subject, HP, (6) evidently under great tension refused to undertake anything of the kind again on the 25th trial. Pegram's subject (3) was not able to score above chance after her run of 25 hits under identical conditions. Riess' subject (9, 10) after scoring 25 straight hits continued to do well for some time, but the difference was that she did not know her scores. It would appear reasonable to conclude, then, that the subject is much affected by knowledge of success in the preceding run. And that, of course, would suggest the wisdom of keeping the subject ignorant of success as Riess did. But it is rare that subjects can be found to work for long without knowing their scores, and a compromise is necessary. There is direct bearing upon the point in discussion from the work of Stuart (11) in which he found that subjects who were not affected by their previous scores were the most successful in ESP tests.

As a result of the comparison of performance in the Formal and Informal Tests, one is driven to the conclusion that to catch the phenomenon of ESP to best advantage a radical change in methodology is needed, a backward one in respect to removing formality. In order, however, to do this without relaxing precautionary features considerable ingenuity will be required. It will be necessary to provide techniques that can be used under conditions that approximate more nearly to play. The subject ordinarily must know, too, from run to run, how well he is doing, and the rules of the experiment should for most subjects be flexible enough to allow each one to follow his own desires quite freely. Obviously this is a large order for the methodologist but it may not be an impossible one.

Summary

A single subject, PH, has been studied through a series of various Informal Tests and a group of four Formal series carried out with the

cooperation of the Parapsychology Laboratory at Duke University. The Informal Tests were strikingly above chance, and the four Formal series, although much lower in score average, gave a combined chi-square which is acceptably significant (.009). Also one of the four was independently significant (CR = 2.8).

The Formal series owe their significance to score totals below expectation; in other words, to negative deviations. This is, however, psychologically as important as a positive deviation, and hypotheses have been offered to account for this depression of scoring below level. The independently reported Stuart-PH Series, which would be classified with the Formal Tests, further substantiates the position taken here.

Comparisons were made as to the success of different p-values on numbers of choices possible in tests. One fifth was suggested as the most successful, but more work is needed for a thorough test.

When success was plotted against key card position, it was observed that V-shaped curves resulted with greatest negative performance on the central key card. This effect was observed in both the open and blind matching methods.

The much more significant scoring in the Informal Tests is contrasted with the low and negative deviations of the Formal, and the differences in conditions are suggested as probable reasons. Chief among these changes is the transfer from a play to a work situation, from freedom and spontaneity to fixed tasks and set rules.

References

1. Foster, A. A. A perception ratio statistic for ESP tests. *J. Parapsychol.*, 1940, **4**, 320-324.
2. Humphrey, B. M., and Pratt, J. G. A comparison of five ESP test procedures. *J. Parapsychol.*, 1941, **5**, 267-292.
3. Reeves, M. P. and Rhine, J. B. Exceptional scores in ESP tests and the conditions. I. The case of Lillian. *J. Parapsychol.*, 1942, **6**, 164-173.
4. Rhine, J. B. ESP tests with enclosed cards. *J. Parapsychol.*, 1938, **2**, 199-216.
5. ———. Experiments bearing upon the precognition hypothesis. III. Mechanically selected cards. *J. Parapsychol.*, 1941, **5**, 1-57.

6. ———. New Frontiers of the Mind. New York: Farrar and Rinehart, 1937.
7. ———. Terminal salience in ESP performance. *J. Parapsychol.*, 1941, **5**, 183-244.
8. RHINE, J. B., PRATT, J. G., STUART, C. E., and SMITH, B.M., with GREENWOOD, J. A. Extra-Sensory Perception After Sixty Years. New York: Henry Holt and Co., 1940.
9. RIESS, B. F. A case of high scores in card guessing at a distance. *J. Parapsychol.*, 1937, **1**, 260-263.
10. ———. Further data from a case of high scores in card guessing. *J. Parapsychol.*, 1939, **3**, 79-84.
11. STUART, C. E. An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *J. Parapsychol.*, 1941, **5**, 99-137.
12. ———. An ESP test with drawings. *J. Parapsychol.*, 1942, **6**, 20-43.

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this journal, the following definitions are submitted for convenient reference. In case of any discrepancy between glossary and usage in the text of an article, the latter should be followed. Words defined elsewhere in the glossary are italicized in the text of the definitions.

AGENT: In tests for *telepathy*, the person whose mental states are to be apprehended by the *percipient*. In *GESP* tests, the person who looks at the *stimulus object*.

AVERAGE SCORE: Average number of *hits* per *run*.

BM (BLIND MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are laid out face-down before him in an unknown order. Unless otherwise stated, the order is also unknown to the experimenter.

BT (BEFORE TOUCHING): The technique in which the top card of the face-down *deck* is *called* and, after being called, is laid aside for *checking* at the end of the *run*. Each card in the *deck* is treated in the same way.

CALL v.: To attempt to identify a *target* or *stimulus object* (or mental state of an *agent* in *telepathy*).

CALL n.: The *response* described above; also the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely *score* if only *chance* obtains.

CHANCE AVERAGE: *Mean chance expectation* in terms of *average* per *run*.

CHECK: To determine a *score* after the completion of a *run* by comparing the order of the *subject's calls* with the order of cards in the deck.

CHI-SQUARE: A sum of quantities each of which is a *deviation* squared divided by an expected value. Also a sum of the squares of *CR's*. (Occasionally the square of a simple CR may be used as chi-square.)

CLAIRVOYANCE: *Extra-sensory perception* of objective events as distinguished from *telepathic* perception (of the mental or subjective events of another person).

COVARIATION: Correlation evaluated in terms of theoretical means and *standard deviations*.

CRITICAL RATIO: The observed *deviation* divided by the *standard deviation.*

DECK: Twenty-five *ESP cards,* five of each suit.

DEVIATION: The amount an observed number of *hits* or an *average score* varies from the *mean chance expectation* or *chance average.* A *deviation* may be total (for a series of *runs*) or average (per *run*).

DT (DOWN THROUGH): The technique in which the cards are *called* down through the *deck* before any are removed or *checked.*

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of *ESP.*

ESP (EXTRA-SENSORY PERCEPTION): Response to an external event (perception) not presented to any known sense.

ESP CARDS: Cards, each bearing one of the following five symbols: star, circle, three parallel wavy lines (called "waves"), square, plus.

ESP SYMBOLS: See plate opposite page 1, this journal, Vol. 1, March 1937.

ESP TESTS: A considerable number of techniques come under this heading which are conveniently represented by initials, the principal ones being: *BT, DT, PT, GESP, BM, OM, STM.*

EXTRA-CHANCE: Not due to *chance* alone.

FREE MATERIAL: *Stimulus objects* that are not limited to a known number of categories.

GESP (GENERAL EXTRA-SENSORY PERCEPTION): A technique designed to test the occurrence of *extra-sensory perception,* permitting either *telepathy* or *clairvoyance* or both to operate.

HIT: The correct correspondence of a *subject's call* or response with a *stimulus card* or *object.*

HIT FREQUENCY DISTRIBUTION: The grouping of the total *hits* in a *series* of *runs* with respect to their original position in the *run.*

KEY CARD: One of the five cards (where there are five suits) against which the cards of the test *deck* (i.e., *target cards*) in the *matching* tests (*OM, BM, STM,* etc.) are *matched.*

MATCHING: A form of *calling* in which a *target card* is placed opposite the *key card* which the *subject* selects to identify it. Also, in the evaluation of *free material,* the act of a judge in identifying a given *response* with a *stimulus object.*

OM (OPEN MATCHING): The technique in which a *subject matches* a *deck* of *ESP cards* to five *key cards* which are face-up before him.

PARAPSYCHOLOGY: A division of psychology dealing with the "extra-normal"—those psychical effects which appear not to fall within the scope of what is at present normal and recognized law.

PERCIPIENT: The person who makes the *calls* in a test situation.

P (PROBABILITY): A mathematical estimate of the expected relative frequency of a given event if chance alone were operative.

RESPONSE: The act of the *subject* in attempting to identify the *stimulus object.*

RUN: A succession of *trials,* usually the *calling* of a *deck* of 25 *ESP cards* or symbols.

RUN SALIENCE RATIO (RSR): A measure of *salience* within the *run.*

SALIENCE: The relation of rate of success in the end *segments* of the *run* to that of the middle *segments;* also the relation of the rate of success in the end *trials* of the *segment* to that of the middle *trials.*
TERMINAL SALIENCE: A higher rate of *deviation* in the end *segments* of the *run* (or in the end *trials* of the *segment*) than in the middle *segments* (or *trials*).
MIDDLE SALIENCE: A higher rate of *deviation* in the middle *segments* of the *run* (or in the middle *trials* of the *segment*) than in the end *segments* (or *trials*).

SALIENCE RATIO (SR): A measure of the relation of the rate of success in the end *segments* of the *run* (or in the end *trials* of the *segment*) and that of the middle *segments* (or *trials*). (For details of the manner of obtaining SR's, see Vol. 5, pp. 193-195.)

SCORE: The number of hits made in one *run.*
TOTAL SCORE: *Score* of any number of *runs.*
AVERAGE SCORE: *Total score* divided by number of *runs.*

SCREEN: An opaque barrier used between the *subject* and the card or *agent.* The main types of screens are illustrated in this journal on their first introduction in print.

SEGMENT: One of the five consecutive sets of five *calls* in a *run* of 25 *trials.* The first five *calls* would constitute the first *segment;* the second five, the second, etc.

SEGMENTAL SALIENCE RATIO (SSR): A measure of *salience* within the *segments* of the *run.*

SERIES: Several *runs* that are grouped in accordance with a stated principle.

SIGNIFICANCE: A numerical result is significant when it equals or surpasses some criterion of degree of chance improbability. Common criteria are: a probability value of .01 or less, or a *deviation* in the expected direction such that the *critical ratio* is 2.5 or greater.

STANDARD DEVIATION: The theoretical root mean square of the *deviations.* For *ESP cards,* $SD = 2\sqrt{\text{no. of } runs}$.

STIMULUS OBJECT: The *ESP card* or drawing or other object, some identifying characteristic of which is to be apprehended by the *subject.*

STM (SCREENED TOUCH MATCHING): The technique in which the *subject* makes his *call* by pointing to one of five positions or exposed *symbols* under a special *screen.* The experimenter places the *target card* so designated in the position pointed to. The *screen* blocks all vision by the *subject* of the *cards* and their manipulation by the experimenter.

SUBJECT: The person who is experimented upon. Most commonly the *percipient* in *ESP,* though also the *agent* in *telepathy.*

TARGET: *Stimulus object.*

TARGET CARD: The *card* which the *percipient* is attempting to perceive (i.e., to identify or otherwise indicate a knowledge of).

TARGET DECK: The *deck* of cards the order of which the *subject* is attempting to identify.

TELEPATHY: *Extra-sensory perception* of the mental activities of another person. It does not include the *clairvoyant* perception of objective events.

TRIAL: A single attempt to identify a *stimulus object.*

THE PARAPSYCHOLOGY LABORATORY OF DUKE UNIVERSITY

The Parapsychology Laboratory is a branch of the Department of Psychology of Duke University. Its beginnings date from 1930 when its research program was launched by members of that department under the sponsorship of Professor William McDougall and with the favorable interest of the President of the University—at that time, Dr. William Preston Few. The first report of the work done was issued in book form in 1934 under the title *Extra-Sensory Perception;* in 1937, the *Journal of Parapsychology* was established; and in 1940, five members of the Laboratory staff collaborated to publish *Extra-Sensory Perception After Sixty Years* which summarized not only the work of the Laboratory but similar research, past and present, in other places as well.

Problems of Research. The main topic of investigation has been extra-sensory perception in its various phases—telepathy, clairvoyance, and precognition. Other work has been planned, and to some extent undertaken; but the main objective has been to arrive first at a fair understanding of extra-sensory perception itself before turning definitely to other problems of parapsychology. However, an experimental staff is being trained and methods developed for a broad and inclusive program which awaits only the time and the necessary financial assistance for its execution.

Financial Support. The Laboratory is supported jointly by Duke University and by a small group of donors who have thus far desired to remain anonymous. The Prince Memorial Fund, established in 1934, provided two fellowships in Parapsychology at Duke over a five-year period. In 1935, the William McDougall Research Fund was created at the University with a special committee, appointed by the major contributor and approved by President Few, to administer it. This is called the Parapsychology Research Committee and is self-perpetuating. In order to insure that the purposes of the contributors will always be respected, the Committee has been given the legal authority to constitute itself a board of trust, independent of the University should the need ever arise.

Thus far, the Laboratory has had no permanent endowment and has been maintained by current contributions. This situation has resulted in a lack of long-term security and has worked definitely against the planning of large undertakings. It has also occasioned the serious loss of valuable, trained personnel.

In the judgment of this Committee, it is highly important that, either by will or gift, grants be made of sufficient size to establish an endowment that would guarantee not only the permanence but the stabilization of the research program of the Parapsychology Laboratory.

(Signed) ALICE M. BALDWIN, *Dean, Woman's College, Duke University (Chairman)*
HORNELL HART, *Professor of Sociology, Duke University*
O. K. MERRITT, *President, Renfro Hosiery Mills*
J. B. RHINE, *Professor of Psychology, Duke University*
The Parapsychology Research Committee, Duke University